Medical insurance and free choice of physician shape patient overtreatment: a laboratory experiment

Huck, Steffen; Lünser, Gabriele K.; Spitzer, Florian; Tyran, Jean-Robert

WZB

Wissenschaftszentrum Berlin
für Sozialforschung

Steffen Huck, Gabriele Lünser,
Florian Spitzer, Jean-Robert Tyran

Medical Insurance and Free Choice of Physician Shape Patient Overtreatment. A Laboratory Experiment

Discussion Paper

SP II 2014–307r

October 2014 (revised June 2016)

Research Area
Markets and Choice

Research Unit
Economics of Change

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Affiliation of the authors:

Steffen Huck, WZB and UCL London
Gabriele Lünser (no academic affiliation)
Florian Spitzer, University of Vienna
Jean-Robert Tyran, University of Vienna

Abstract

Medical Insurance and Free Choice of Physician Shape Patient Overtreatment. A Laboratory Experiment*

In a laboratory experiment designed to capture key aspects of the interaction between physicians and patients, we study the effects of medical insurance and competition in the guise of free choice of physician, including observability of physicians' market shares. Medical treatment is an example of a credence good: only the physician knows the appropriate treatment, the patient does not. Even after a consultation, the patient is not sure whether he received the right treatment or whether he was perhaps overtreated. We find that with insurance, moral hazard looms on both sides of the market: patients consult more often and physicians overtreat more often than in the baseline condition. Competition decreases overtreatment compared to the baseline and patients therefore consult more often. When the two institutions are combined, competition is found to partially offset the adverse effects of insurance: most patients seek treatment, but overtreatment is moderated.

Keywords: Credence good; Physician; Overtreatment; Competition; Insurance.

JEL classification: C91, I11, I13.

---

* Corresponding author: Jean-Robert Tyran, University of Vienna, Vienna Center for Experimental Economics (VCEE), Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria; jean-robert.tyran@univie.ac.at. Huck is at WZB Berlin and UCL in London; steffen.huck@wzb.eu. Lünser has no academic affiliation. Spitzer is at the Department of Economics at the Vienna University of Economics and Business and at the Vienna Center for Experimental Economics (VCEE); florian.spitzer@wu.ac.at. Tyran is also affiliated at the Department of Economics, University of Vienna, at the Department of Economics, University of Copenhagen, and at the Centre for Economic Policy Research (CEPR), London. We gratefully acknowledge financial support from the Austrian Science Fund (FWF) under project I 2027-G16.

1. Introduction

We study the economic incentives emanating from two key institutions in the medical market – competition and insurance – conceptualizing interaction as the provision of a credence good. Markets for credence goods are characterized by a high degree of asymmetric information between those supplying and those demanding the good or service. Medical treatments are a prime example of credence goods, and an economically important one.[1]

The specific interaction we study is as follows. A patient is confronted with a medical problem and chooses whether to consult a physician. The medical problem can be severe in which case only a severe (and costly) treatment can provide a cure. Alternatively, the problem could be mild, such that a mild (and cheap) treatment is sufficient for a cure. Information about the type of problem is asymmetric: after examining the patient the physician knows what kind of treatment the patient needs, but the patient does not. We induce incentives for overtreatment (that is, to provide the severe treatment when the problem is in fact mild) by choosing experimental parameters such that the physician makes more money from the severe treatment. Reputational incentives disciplining physicians are weak because the patient only learns that he has been cured, but not whether the treatment he received was appropriate.[2] Such markets are likely to be beset by overtreatment and low efficiency.[3]

We study how basic forms of medical insurance and competition shape overtreatment and other outcomes in this setting. We study competition in the guise of patients being able to freely choose among physicians.[4] This type of competition has been shown to be rather effective in markets for experience goods (Huck et al. 2012). Competition has bite in such markets because reputational incentives are strong and can discipline sellers to provide good quality. But with credence goods, building effective reputations is difficult because patients cannot tell whether a severe treatment was necessary. Nevertheless, we find that competition

[1] Other examples of credence goods are car repairs (e.g. Schneider 2012) or taxi rides in a foreign city (e.g. Balafoutas et al. 2014). Both studies provide field experimental evidence. For an overview of the theoretical literature, see Dulleck and Kerschbamer (2006). The seminal paper on markets for credence goods is Darby and Karni (1973).

[2] A key difference between credence and experience goods is that overcharging or overprovision cannot easily be detected (see Dulleck et al. 2011 for a discussion).

[3] Iizuka (2007) for instance reports evidence from the Japanese prescription drug market where physicians do not only prescribe but also dispense drugs. They show that prescriptions are to some extent influenced by mark-ups and hence not only by factors that are relevant to the patient's state of health.

[4] Note that this type of non-price competition is typical for patient-physician interactions in which prices are regulated. See Huck et al. (2016) for an experimental study of price competition in a market for experience goods. We use "competition" and "free choice of physician" interchangeably in the remainder of the paper.

has surprisingly strong beneficial effects. It clamps down on overtreatment (the incidence falls by about two thirds) encouraging patients to consult more often as they can now be more confident not to be overtreated.

The second institution we investigate is insurance. We expect insurance to invite moral hazard, as it shields the individual patient from the adverse monetary consequences of overtreatment. The insurance we study socializes the cost of overtreatment. As physicians anticipate or become aware that patients are less wary under the umbrella of insurance, they have an additional incentive to overtreat. We expect reduced wariness to mitigate the disciplining effect of reputational concerns. Indeed, this is what we find: the consultation rate is much higher with insurance than in the baseline, and overtreatment is more common as a consequence.

By virtue of a 2-by-2 design we can also study interaction effects. We find that competition has powerful effects both in the absence and in the presence of insurance. In the latter case, competition cuts overtreatment in half and boosts the share of consulting patients. Thus, competition partly mitigates the adverse effect of insurance while keeping incentives to consult strong. As a result, the combination of both institutions produces the highest level of public health among the institutional settings studied here. This combination is however also associated with the highest expenditures for health (measured by the total transfer from patients to physicians). At least in the setting studied here, it does not seem possible to decrease expenditures without decreasing public health at the same time.

In our experimental design, competition involves observability of market shares. Patients do not only have the possibility to choose the physician they want to interact with but they can also observe by how many other patients a physician has been consulted in the past. Hence, we can only draw conclusions about the joint impact of competition and observability of market shares. To be able to disentangle these two effects, we conduct two additional control treatments in which patients are able to choose their physician but market shares cannot be observed. The results of these control treatments indicate that the positive impact of competition is mainly driven by free choice and not by observability. In the remainder of the paper, we focus on our four main treatment conditions[5] and describe the additional treatments as well as their results in Sections 4.3 and 4.4.

[5] We will use the expression "impact of competition" as a synonym for the impact of competition in the guise of "informed choice," i.e. for the joint impact of being able to choose a physician ("pure competition") and the observability of market shares. The primary reason to focus on the main treatment conditions is to keep the structure of the paper clear and concise. Moreover, it seems sensible to vary the observability of market shares

We think our results speak to ongoing debates about how to devise efficient systems in health care. Free choice of physician and the availability of medical insurance are among the most relevant institutional choices to make in the design of a health care system. For example, there is an ongoing debate in various countries whether elements of co-payment should be increased to overcome moral hazard problems associated with health insurance. Health care systems also strongly differ by the degree to which patients are allowed to choose their physician: With a general practitioner-centered model, patients are usually assigned to a physician in their district and possibilities to consult different physicians are restricted – in contrast to health care systems with free choice of physician. We think that our study sheds new light on these important debates by virtue of the ability to measure and control important aspects of the patient-physician interaction. For example, we unambiguously observe all instances of overtreatment and we control the cost it entails. In the field, overtreatment often goes unnoticed and its costs can only be roughly estimated. Our treatment variations also allow us to isolate the effects of institutional changes to a much higher degree than is possible in the field. However, circumspection is advised in extrapolating from our highly stylized setting to the actual policy debate which is embedded in a rich medical-technical, institutional, and cultural context. Such context-specific aspects may or may not matter for the interaction of patient and physician. What we provide here is an analysis of how economic incentives emanating from controlled but highly stylized institutional changes shape overtreatment in an environment that is conducive to it.

Related literature. Our study is related to various streams of literature. First, it contributes to the recently emerging literature in experimental health economics. A series of laboratory experiments (Brosig-Koch et al. 2013a, 2013b, Hennig-Schmidt et al. 2011, Kairies and Krieger 2013, Keser et al. 2014, Keser et al. 2013, Green 2014) investigate incentive effects of remuneration systems for physician behavior. For example, Hennig-Schmidt et al. (2011) compare a capitation system (in which the physician gets paid per patient independent of the treatment provided) and a fee-for-service system (in which payment does depend on the treatment provided). They find that subjects react to the incentives of the payment system – leading to substantial levels of under- and overtreatment – and that this is also the case for medical students. In contrast to our study, which focuses on patient-physician interaction, patients make no choices in their experiment (their payoffs are modeled by donations to a medical charity that uses the money for medical treatment of real patients). Their finding that

---

and pure competition at the same time. Patients can only react to the market shares they observeif they are able to choose a physician in the first place. The presence of pure competition should therefore promote the emergence of institutions facilitating the observability of markets shares.

financial incentives shape the treatment physicians provide is also supported by various empirical studies using field data (see e.g. Clemens and Gottlieb 2014, Devlin and Sarma 2008 or Sørensen and Grytten 2003). Another example for a health economic experiment is Schram and Sonnemans (2011) who study the demand side of a health insurance market, i.e. how subjects choose a health insurance policy in a complex decision environment. Buckley et al. (2012) investigate the interplay of public and private health insurance in a revealed-choice experiment.

Our experiment is also related to a stream of experimental literature investigating credence goods, in particular Beck et al. (2013, 2014) and Mimra et al. (2013, 2014). A close match to our study is Dulleck et al. (2011). These authors study a market for credence goods in a flexible and broad setting that allows them to analyze various institutional frameworks and various aspects of market failure in the provision of credence goods. For example, they allow for overtreatment (as we do), in addition to overcharging and undertreatment. These phenomena are particularly relevant in markets for car repairs but less characteristic of many markets for medical treatments. We therefore focus on a framework with fixed prices and overtreatment that fits the patient-physician interaction and allows us to study relevant institutions like health insurance and free choice of physician. Mimra et al. (2013) is also closely related to our paper. These authors study the effect of price competition compared to fixed prices. They find that the level of supplier opportunism (in their case undertreatment and overcharging) is significantly higher in a market with price competition than in a market with fixed prices. In contrast to their study, we compare competition with fixed prices to a situation with fixed assignment (i.e. random repeated matching) which is particularly relevant in a market for medical treatments.

There are only a few experimental studies investigating the effects of free choice of interaction partner based on reputation. Huck et al. (2012, 2016) and Bolton et al. (2008) study free choice of seller in a market for experience goods, Dulleck et al. (2011) and Mimra et al. (2013) in a market for credence goods (in their setting, competition is based both on reputation and prices). The main finding of these studies is that competition with fixed prices decreases opportunistic seller behavior whereas price competition pushes prices down but increases opportunism at the same time. In the health context, several empirical studies suggest that free choice of the health-care provider has beneficial effects on market performance. For example, Cooper et al. (2011) find that a reform in the English National Health Service reduced mortality significantly by giving patients the freedom of choice which hospital they want to be transferred to. Kalda et al. (2003) and Schmittdiel et al. (1997) find that giving patients the option to choose the physician providing primary care leads to higher overall patient satisfaction.

A series of empirical studies estimate the extent to which the demand for medical services is related to the extent of insurance coverage, e.g. the co-payment rate. Examples are Scitovsky and Snyder (1972), Manning et al. (1987), and Aron-Dine et al. (2013). Most of these studies find that increased co-payment reduces the demand for medical services. Chiappori et al. (1998) provide a particularly convincing study on this matter. These authors analyze data from a natural experiment where a co-payment rate of 10% was introduced for one group of patients but not for a control group of patients. They find that the number of home visits decreases significantly with the co-payment rate but find no effect for the number of office visits. Sülzle and Wambach (2005) provide a theoretical analysis of insurance in a market for credence goods with the possibility to search for second opinions. They show that a higher rate of co-insurance can have two opposite effects. It can either lead to less fraud and less search for second opinions or to more fraud and more search activities in the market.

While the evidence above for insurance-induced moral hazard is rather abundant and quite compelling, we are not aware of evidence on supply-side responses to such moral hazard (i.e. to what extent physicians provide more services than necessary if they anticipate that patients care less about getting excessive treatments because the costs are covered by insurance). Such responses could be called “second-order moral hazard” and Balafoutas et al. (2013) provide evidence for it in the context of taxi rides in Athens. They find that if a passenger indicates to the driver that the bill is paid by their employer, passengers are significantly more likely to be overcharged compared to a control group giving no such indication.

We proceed as follows. Section 2 describes the experimental design, section 3 derives predictions for the effects of competition and insurance, section 4 presents results, and section 5 concludes. Appendix A provides instructions, B screenshots, and appendices C to H provide complementary tables, figures and analyses.

2. Experimental Design

Before going into a detailed description of treatment conditions, parameters and procedures, we now provide a short overview of the design.

In all conditions, experimental subjects are randomly assigned to a fixed role as physician (the seller or provider of the treatment who is called “adviser” in the experiment) or patient (the buyer or demander of the treatment who is called “client” in the experiment) at the beginning of the experiment and they interact repeatedly. Patients know that they have a problem (mild or severe) and need a treatment, but they do not know what treatment they need. In contrast, physicians do know what type of treatment the patients need. Patients choose whether or not to consult a physician. The material incentives in our experiment are stacked against providing the correct treatment when the patients need a mild treatment, i.e. physicians have strong incentives to overtreat patients. Reputational incentives to provide proper treatment are weak because the patients cannot unambiguously infer whether they got the treatment they needed or whether they were overtreated.

Starting from this baseline condition, we investigate the effects of two simple institutions, competition and insurance, on overtreatment and other interaction outcomes (i.e. consulting rates, patients' and physicians' average earnings as well as two measures for efficiency). Competition means that patients can choose which physician they want to consult rather than being assigned randomly to a physician. Insurance means that the cost of treatment (or more precisely the additional cost of a severe treatment) is borne by all patients collectively rather than by one patient alone.

Table 1: Main Treatments

		Insurance	
		No	Yes
Competition	No	**BASE** *n* = 56	**INS** *n* = 56
	Yes	**COMP** *n* = 56	**INS-COMP** *n* = 56

Notes: We have 7 markets per main treatment. In each market, 5 patients and 3 physicians interact. The number of subjects in the main treatments is 224 (= 4 treatments x 7 markets x 8 participants). We have another 112 subjects in 2 control treatments described in section 4.3.

Table 1 summarizes the design. We use a between-subject 2 x 2 factorial design and label the main treatments as follows: BASE (baseline condition), COMP (competition but no insurance), INS (insurance but no competition) and INS-COMP (insurance and competition). We also ran two additional control treatments described in section 4.3.[6]

2.1 Baseline condition

We consider the interaction of three physicians and five patients in a market (i.e. a matching group).[7] At the beginning of the experiment, subjects are randomly assigned to a role and market for the entire experiment which consists of 30 periods. At the beginning of each period, patients are randomly assigned to a physician in their market. Thus, each physician may find herself with between 0 and 5 patients assigned to her, and all or some of the patients assigned to her may also consult her.[8]

Figure 1 shows the stage game between one physician and one patient who has been matched to her. The structure of moves and the payoffs are common information to all participants. At the beginning of each period, the severity of the patient's problems is randomly determined (same draw for all patients). It is mild with probability $q(M)$ and severe with probability $q(S) = 1 - q(M)$. When patients make the choice whether to consult (C) or not to consult ($\neg C$), they are not aware of the severity of their problem (indicated by the information set marked with "Patient" in figure 1).

In contrast, the physicians do know the severity of the patients' problem and the number of patients consulting them.[9] The physician then chooses the treatment (m or s) for the patients who have consulted her. Given a mild problem, the physician chooses whether to overtreat the patient, i.e. she has the option to provide a severe treatment (choose s when the Problem is M, see left node marked with "Physician"). In case of a severe problem the physician cannot undertreat (e.g. decline to treat the patient). That is, in case of a severe problem, she has to provide the severe treatment s.[10]

---

[6] To make the choices among suppliers effective, we provide information about market shares in COMP which is absent in BASE. That is, the effect of BASE vs. COMP (and INS vs. INS-COMP) is driven by "informed choice". We have run additional control treatments to disentangle the effect of (uninformed) choice of physician and of providing information about market shares alone, see section 4.3.

[7] Physicians are called "advisers" and patients are called "clients" in the experiment, see section 2.3 for explanations.

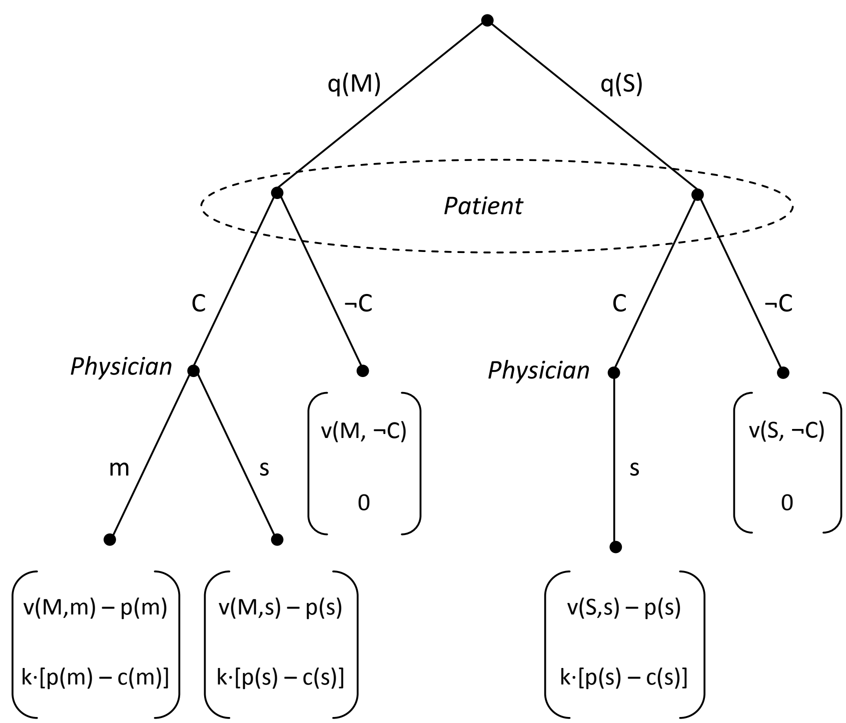
[8] We use female gender for physicians and male gender for patients throughout to facilitate understanding.

[9] Physicians also know the number of patients assigned to them but do not know the identity of the patients.

[10] The physician has to provide the same treatment to all patients who consulted her. The reason for assigning the same type of problem to all patients within one group is to make sure that the physician has always the possibility to provide the proper treatment to all consulting patients.

The payoff earned by a physician from interacting with one patient is shown in the last line of figure 1. The payoff is the price of treatment (which is assumed to be exogenously fixed and has to be paid by the patient to the physician) minus the cost of treatment: $p(i) - c(i)$, $i \in \{m, s\}$. The total payoff for one period is this number multiplied with the number of patients k who consulted the physician. Accordingly, the final payoff is zero if no patient consulted the physician. Note that the physician's payoff results from actually treating the patient; just being matched generates no value for the physician.

Figure 1: Baseline condition (stage game)



The payoff earned by a patient depends on whether he decided to consult the physician or not. If the patient decides not to consult, his payoff depends on the severity of the "unsolved" problem. We make the rather plausible assumption that the patient's payoff is lower when his unsolved problem is severe than when it is mild, i.e. $v(S, \neg C) < v(M, \neg C)$.

If a patient consults a physician, he will always receive a treatment so that his problem is solved for sure. The payoff of the patient is determined by the value of the treatment (which depends (for now) on the severity of the problem as well as the treatment provided) and the price of the treatment: $v(j, i) - p(i),\ i \in \{m, s\},\ j \in \{M, S\}$.

The reputational incentives to mitigate overtreatment are rather weak given the feedback provided to patients. At the end of each period, the patient is informed about the treatment he got but not about the severity of his problem. He of course only learns about the treatment he got given that he consulted the assigned physician. In case the patient does not consult he gets to know the severity of the problem.[11]

When making choices, subjects see a history table showing a summary of previous periods (see appendix B for screenshots). Physicians have fixed IDs which are revealed to patients, i.e. physicians are not anonymous to patients. In fact, when making the consulting choice, patients can see which physician they have been assigned to, whether they had consulted the assigned physician before and what treatment they had gotten from this physician (but do not learn the true severity of the problem they had). For periods in which the physician was not consulted, patients can see the severity of their problem (following the logic explained in footnote 10). When making the choice of what treatment to provide, physicians see the number of patients assigned to themselves, the severity of the patients' problems in the current period and can review the same information for earlier periods, including what treatments they provided.

In essence, the information provided in the history table means that patients can recall their own experiences with a physician but do not know about the experiences of other patients (or the treatments provided by non-consulted physicians). This seems plausible in the context of the interaction between patient and physician.[12]

A characteristic feature of a credence good is that some quality uncertainty persists even after the purchase of the good. We study the type of credence good where the consumer does not know what he needs but can observe what he got.[13] This type of credence good is particularly relevant for medical treatments: The patient can typically observe the treatment he

---

[11] The reason for informing patients about the severity of their problem (after not consulting) is the following: If the patient decides not to consult, he does not receive a treatment and hence his problem is not solved. This implies that the patient experiences the consequences of his unsolved problem (e.g. suffers pain). But he suffers more in case of a severe problem (leading to a lower payoff). Therefore, non-consulting patients can infer the severity of the problem from their payoff (which they learn at the end of each period). In contrast, the problem is solved (i.e. he is cured) if the patient consults a physician and receives a treatment. He can therefore not infer whether the problem was severe or mild.

[12] The information conditions here parallel the treatment with private information (pi-nc) in Huck et al. (2012).

[13] The literature (see Dulleck et al. 2011 for a discussion) distinguishes between this type of credence good and a second type (clients know what they want but not what they got). The second type refers to goods where consumers have strong preferences over certain characteristics of a product (like environmentally friendly production) that can however not easily be observed after the purchase.

received from his physician but he is uncertain about his health condition, i.e. what kind of treatment he needs. To parallel the logic in the field context, it is therefore important that the patient cannot determine ex post whether a severe treatment was actually necessary due to a severe problem or whether he was overtreated (i.e. a mild treatment would have been sufficient). To guarantee that this is the case, the payoff from a severe treatment needs to be independent of the actual severity of the problem (otherwise the patient could easily infer whether the severe treatment was necessary or not):

$$v(M,s) - p(s) = v(S,s) - p(s) \quad \Leftrightarrow \quad v(M,s) = v(S,s)$$

Furthermore, we choose the parameters such that the sum of the patient's and physician's payoff is independent of the treatment provided by the physician:

$$v(M,m) - p(m) + p(m) - c(m) = v(M,s) - p(s) + p(s) - c(s)$$

This choice enhances our experimental control as it allows us to exclude a concern for efficiency as a motive for the physician's choice of treatment. As a consequence of our parameter choices, overtreatment (i.e. providing a severe treatment in case of a mild problem) is not associated with an efficiency loss; it is a pure redistribution from the patient to the physician.[14]

We choose parameters such that the appropriate treatment in case of a mild problem (i.e. the mild treatment) generates at least as much value as the inappropriate (severe) treatment, i.e. $v(M,m) \geq v(M,s)$, and that the cost of a severe treatment are at least as high as the cost for a mild treatment, i.e. $c(s) \geq c(m)$. Given these choices, it follows that $v(M,s) = v(M,m)$ and $c(s) = c(m)$.[15] Essentially, the fact that $v(M,s) = v(M,m) = v(S,s)$ means that receiving a treatment solves the medical problem (i.e. the patient is cured), and this is independent of whether the problem was severe or mild. Thus, we assume that there are no adverse health effects from being overtreated.

14 This choice is not motivated by parallelism to medical practice (overtreatment may for instance cause inefficiency in the presence of capacity constraints for severe treatments) but for methodological reasons. The absence of efficiency losses allows us to implement a clear-cut case of a credence good: payoffs after being overtreated and after receiving a necessary severe treatment are identical. Therefore, patients cannot identify from inspection of their payoffs whether they received the correct treatment or not. Since we hold this property constant across all conditions it does not affect the interpretation of treatment comparisons.

15 It would not be correct to infer from this that overpricing is entirely isomorphic to overtreatment, the case we discuss here. The reason is that we assume that only a severe treatment can solve a severe problem, indicating that a severe and a mild treatment differ not only in price.

2.2 Treatments with Insurance and Competition

The insurance condition is identical to the baseline condition except for patients' payoffs. In the baseline condition, a consulting patient's payoff depends only on the treatment he got. The downside of being overtreated results from (unnecessarily) having to pay a higher price. In the insurance conditions, the patient is shielded from (i.e. insured against) incurring the additional cost of being overtreated. Specifically, the additional costs of a severe treatment are socialized in the sense that they are borne by all patients collectively rather than by the overtreated patient alone.

All patients pay an insurance premium to cover the additional costs of overtreatment; this premium depends on the total number of severe treatments within a market $n(s)$ and it is used to pay the price difference between a mild and a severe treatment ($p(s) - p(m)$). Note that patients who do not consult a physician also pay this premium. The premium is therefore the total additional spending for severe treatments divided by the total number of patients in one market (N):

$$P(n(s)) = \frac{n(s)}{N} \cdot \big(p(s) - p(m)\big)$$

A patient's payoff for refraining from consulting is $v(S, \neg C) - P(n(s))$ for a severe problem and $v(M, \neg C) - P(n(s))$ for a mild problem. If the patient decides to consult a physician, his payoff is $v(j, i) - p(m) - P(n(s))$, $i \in \{m, s\}$, $j \in \{M, S\}$. While this expression can turn negative, we cap patients' payoffs at zero to prevent loss aversion to shape behavior.

The calculation of payoffs for physicians is identical to the baseline condition. As explained above, $v(M, s) = v(M, m) = v(S, s)$. This means that the individual payoff of a patient does – in contrast to the baseline condition – no longer depend on which treatment he gets. The only effect of overtreatment is that it boosts the insurance premium which has to be paid by all patients in the market collectively (also those who did not consult a physician).

Because patients are informed about how the insurance premium is calculated and learn the premium they have to pay at the end of every period in INS, patients do not only get to know the treatment they received themselves (as in BASE) but can also infer the total number of severe treatments within the market from their final payoff. However, as in BASE, they are not informed about the true severity of their problem (or the severity of the problem of other patients). Note that the insurance premium is calculated to be fair (covers all costs but does not generate surplus).

The treatments with competition (COMP and INS-COMP) differ from those without competition (BASE and INS, respectively) in two ways. The first is that patients now can choose freely which physician to consult. The matching of patients and physicians is thus not random as in BASE (and INS) but endogenous. In treatments with competition, patients do not only decide whether to consult a physician but also which one to consult. The second difference concerns information. With competition (i.e. in COMP and INS-COMP), both patients and physicians see the market shares, i.e. the number of patients having consulted a particular physician in previous periods in the history table. The effect of treatment variation in the main treatments (i.e. BASE vs. COMP and INS vs. INS-COMP) thus measures the effect of introducing what we will call “informed choice” or “informed competition” below. Making such information available quite plausibly boosts the effect of physician choice, and vice versa. Because we need to limit the duration of the experiment for practical reasons (e.g. to prevent subject fatigue), it is difficult for physicians to effectively form reputations and for patients to reliably infer them within the time of interaction available (30 periods). Access to such information in the field seems plausible. Patients can often observe whether a physician is in high demand (they can e.g. observe the length of the waiting list or how full the waiting room is). While we think the effect of “informed choice” is highly relevant, we also ran two control treatments to isolate the effect of choice of physician when market shares cannot be observed (COMP_nms and INS-COMP_nms, see section 4.3).

2.3 Experimental procedures and parameters

The experiment was conducted using the software z-tree (Fischbacher 2007) with a total of 336 undergraduate students from various disciplines at the University of Copenhagen as subjects (224 in the main treatments and 112 subjects in the two control treatments, see section 4.3)[16]. Subjects were recruited using the online recruiting system ORSEE (Greiner 2015) and each subject participated in one session only. At the beginning of the experiment, subjects were seated randomly in the laboratory and received written instructions (see appendix A) explaining the experiment. The language of the instructions was kept neutral. We did not frame the situation in a medical context. Instead of “physician” and “patient”, we used the terms “adviser” and “client”, and explained that the latter was confronted with a problem that could either be mild or severe.

[16] Henning Schmidt and Wiesen (2014) show that medical students behave more pro-socially than students from other fields when a game broadly akin to ours is framed in terms of provision of medical treatments.

In a laboratory experiments, such framing may be important. Previous studies find that changing a single word (Dufwenberg et al. 2011) or the labelling of actors (Huck et al. 2004) can result in a remarkable difference in behavior. Engel and Rand (2014) find that subjects might project their own frame when they are confronted with a decontextualized decision situation, leading to differences in behavior if the background of the experiment does not coincide with the projected frames. We choose the somewhat colorful terms “problem”, “adviser”, and “client” to find a compromise between the particularly loaded medical framing on the one hand and abstract expressions like “A-participant” and “B-participant” on the other hand. As an alternative to abstract expressions, the main purpose is to facilitate subjects’ understanding of instructions in a rather complex set-up. We have decided not to use a medical framing for the following reasons: First, it allows us to interpret our findings also in the context of other market for credence goods (like car repair services), so that we are able to contribute to the experimental literature on markets for credence goods as well (where the instructions generally use a neutral framing, see for instance Dulleck et al. 2011). Second, the medical framing seems more important in situations where undertreatment can be chosen by suppliers and where the consequences for a patient might be suffering or even death. Such outcomes are, of course, much harder to capture through purely monetary incentives. The focus of our study is, however, on overtreatment (think of a radiologist deciding on whether to employ a harmless, expensive and wholly unnecessary scanning procedure to increase his payout). The decision to overtreat a patient with a harmless procedure is in our view mainly an economic rather than a medical decision. After all, the patient is cured with or without overtreatment.

Moreover, we are mainly interested in treatment differences rather than absolute levels of behavior. Assuming that there is no interaction effect between the labelling of actors in the instructions and the analyzed institutions, framing should only play a minor role. This assumption is supported by the fact that none of the papers investigating the impact of a medical vs. a non-medical framing (see for instance Ahlert et al. 2012, Boehm et al. 2015) finds a notable interaction effect between the framing and other treatment variations.[17]

Our experimental design differs in one other important dimension from previous economic experiments investigating how physicians treat their patients. In these studies (Hennig-Schmidt et al. 2011 is the first study implementing this method), patients do not play an active role in the experiment, as the level of treatment provided by the physician is modeled

[17] One exception is Kesternich et al. (2015) who find that a Hippocratic Oath is more effective with the medical framing. However, this treatment variation is directly related to the medical framing, which is not the case for our treatment variations.

as a contribution to a charity with a medical background (i.e. the higher the level of treatment chosen by the physician the higher the resulting donation to the charity and, as a consequence,the level of treatment received by real patients taken care of by the charity). This approach has certainly the advantage that physicians' actions do influence the well-being of actual patients which (in combination with the medical framing) increases ecological and external validity.

A limitation of this method is that patients have no active role in the experiment. An actual interaction between patient and physician is therefore not possible. This is not much of an issue in the studies mentioned above, as their focus is on physicians' behavioral responses toward changes in the incentive structure of the remuneration system. In contrast, one of the main goals of our experiment is to study competition in the guise of free choice of physician in a market for medical services. This requires actual interaction between physicians and patients and to model the latter as active participants in the laboratory.

On average, one session lasted about 75 minutes and subjects earned 212 DKK (≈ 28.4 EUR), including a show-up fee of 75 DKK. The severity of the patients' problem was determined randomly (with overall proportions of 1/3 and 2/3) for each market and period in the baseline condition in preparation of the experiment. The same sequences were then also used for the markets in the other conditions (i.e. each market in a given condition had the same order of periods with mild and severe treatment as another market in the other three conditions).

Figure 2 shows the parameters: The probability of a mild problem is $q(M) = \frac{2}{3}$, and that of a severe problem is $q(S) = \frac{1}{3}$.[18] The patient's benefit if his problem is solved is $v(M,s) = v(M,m) = v(S,s) = 25$.[19] Note that this benefit is independent of the treatment and the severity of the problem. The costs of providing a severe treatment (which are identical to the costs of providing a mild treatment) are $c(s) = c(m) = 5$. The price (to be paid by the patient to the physician) is $p(m) = 15$ for a mild treatment, and $p(s) = 22$ for a severe treatment.

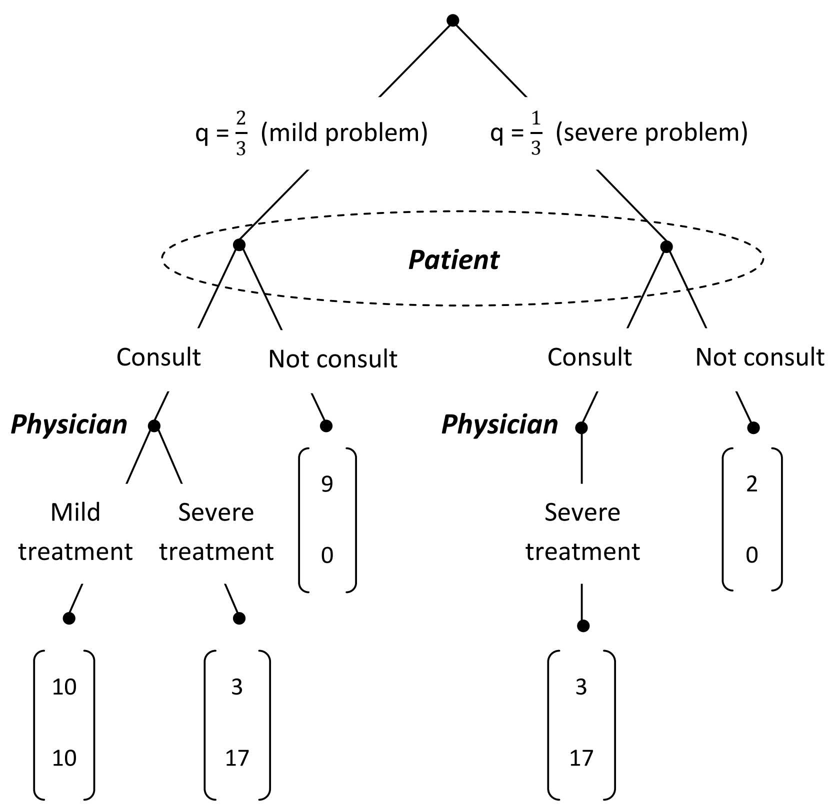
[18] We chose the mild problem to occur twice as often as the severe problem because only periods with a mild problem are interesting with respect to overtreatment. While the periods with a severe problem are not interesting in themselves (physicians are forced to provide the severe treatment and therefore have no real choice to make), they are an essential element of the design to maintain patients' ex-post uncertainty about whether the treatment was necessary.

[19] The payoffs in the experiment are determined by the difference between price and cost (for physicians) and the difference between the value of treatment and the price (for patients). The suggested parameter values are for illustration purposes only; adding a constant to all parameter values would lead to the same payoffs in the experiment.

Hence, the physician's payoff for treating one patient is 17 in case of a severe treatment and 10 in case of a mild treatment.

Her final payoff for one period results from multiplying this number with the total number of patients k consulting her. The patient's payoff from receiving a severe treatment is 3 (and 10 for a mild treatment, respectively). In case the patient does not consult the physician, his payoff is $v(S, \neg C) = 2$ in case of a severe problem and $v(M, \neg C) = 9$ in case of a mild problem. Note that our choice of parameters implies that the patient's benefit of receiving the correct treatment vs. no treatment is small (1 point) while his cost (and the physician's incentive to overtreat) is rather big. This choice of parameters has the advantage that a positive effect of competition, if any, is underestimated rather than overestimated. Our choice of parameters is also driven by the concern to provide an informative baseline condition. To be able to detect any (beneficial and harmful) effects of the institutions under consideration, the baseline needs to be calibrated such that consulting and overtreatment are at intermediate levels, i.e. that there is sufficient room for improvement and for deterioration.

Figure 2: Extensive form (baseline condition, with actual payoffs)



3. Predictions and hypotheses

This section derives theoretical predictions and formulates hypotheses regarding expected treatment differences. We derive equilibria in the stage game for the baseline condition (section 3.1), the conditions with insurance (3.2), and we discuss the condition with competition (3.3). Based on the results in previous studies, we expect the treatment differences to be smaller than predicted by standard theory but to be qualitatively in line with standard theory.

3.1 Predictions for the baseline condition (BASE)

If we assume common knowledge of rationality and strict self-interest, we can solve the game in the baseline condition (see figure 2) by backward induction. Given that the physician will always provide the severe treatment, the patient's expected payoff for consulting is always lower than the expected payoff of not consulting. In the unique equilibrium[20] patients do therefore not consult (and physicians provide a severe treatment if they get the chance to do so – which they do not along the equilibrium path).

From a behavioral perspective, we expect the consulting rate in the baseline condition to be low, but not to be zero as predicted under the standard assumptions. In fact, the experimental evidence from repeated trust games (see e.g. Bolton et al. 2004) reveals a substantial level of trust in initial periods which eventually fades. Yet, two aspects of our design lead us to expect lower consulting rates compared to previous experiments on repeated trust games. First, in markets for credence goods it is much more difficult to build reputations than in markets for experience goods (trust games can be thought of stylized representations of such markets). Thus, patients find it very hard to learn whether they received the treatment they needed from a particular physician (i.e. whether trust was honored or not). Second, our parameter choices imply that the incentives are stacked against consulting the physician: the payoff from receiving the correct treatment is only one point higher than the outside option, whereas being overtreated reduces the payoff by seven points (see figure 2).

As pointed out in the description of the equilibria, physicians have an incentive to provide the severe treatment whenever they get a chance. However, we expect the overtreatment rate to be below 1 as we know from many previous experiments (e.g. Bohnet et al. 2005, Riedl and

[20] The baseline game is a dynamic game with incomplete information. In a strict sense, the appropriate equilibrium concept is the Perfect Bayesian equilibrium. However, as there is no action of another player involved when patients form their beliefs (about their own unknown type), the belief has to be equal to the underlying probability distribution. Moreover, the physician is informed about the patient's type when deciding on the treatment, so applying backward induction is appropriate to solve the game.

Tyran 2005) that many subjects tend to honor trust (i.e. reciprocate) even when this is not in their own monetary interest.

3.2 Predictions in the insurance condition

In the conditions with insurance, the (additional) cost of an unnecessary severe treatment is socialized, i.e. shared among all patients in the market. Intuitively, this means that patients have little incentive to avoid being overtreated. In INS, this creates incentives for patients to consult the physician they have been randomly matched with irrespective of whether she is in good standing (moral hazard). In INS-COMP, insurance undermines incentives to be choosy in whom to consult compared to the case without insurance (COMP). Thus, in both cases, insurance is expected to create incentives for consulting. The unique equilibrium in which patients consult the assigned physician and physicians always provide a severe treatment is formally derived in appendix E.

Behaviorally, insurance is expected to increase the consulting rate and the overtreatment rate but not to the level predicted by standard theory (i.e. we expect consulting and overtreatment to be higher in INS compared to BASE and in INS-COMP compared to COMP, respectively). Note that in the conditions with insurance, a patient's decision of whether to consult a physician (given he expects to be overtreated) is equivalent to the choice faced in a public good game: The individual benefit of receiving the treatment exceeds the disadvantage of having to pay a higher insurance premium. As a group however, patients would be better off if no patient consulted a physician (i.e. individual rationality contradicts collective rationality). The experimental literature on linear public good games (for an overview, see Ledyard 1995) shows that subjects contribute (to some extent) to a public good even though it is not individually rational. For this reason we do not expect insurance to increase the consulting rate to the level predicted by standard theory.

3.3 Predictions for conditions with competition

In the treatment conditions allowing for competition, patients can freely choose between the physicians in their market. As in the conditions without competition, the history table provides each patient with a summary of what kind of treatment he himself got (but not the severity of the problem). But in addition to what patients in treatments without competition see, patients in COMP and INS-COMP also get information about how many patients visited each physician, i.e. the market shares of each physician, but not what treatment the other patients got. This additional information becomes increasingly useful in gauging the trustworthiness of physicians as the experiment progresses. When the proportion of severe treatments provided by a particular physician deviates too strongly from the base rate (1/3) of a severe treatment, patients can estimate that this physician was likely to have overtreated. Physicians now have a reputational incentive to provide the required treatment if patients systematically choose to consult the physician with the best odds to have treated the patient correctly. This reputation mechanism is amplified by the fact that patients can observe other patients' past consulting choices. Patients can now choose physicians with whom other patients apparently have made good experiences (indicated by a high number of visitors).

While information about market shares has some potential to create reputational incentives, these incentives build up only slowly with experience and are likely to remain weak even towards the end of the game because the inference problem is difficult with a noisy signal. When a patient gets a severe treatment, he is uncertain whether this treatment was appropriate or excessive.[21] Another limit to the power of reputational incentives is the fact that patients can only recall their own experience with a given physician (if they consulted her at all) but not how this physician treated other patients in the market. A patient can therefore only slowly benefit from other patients' experiences by shunning physicians with a decreasing market share. Despite these two limitations of patient's possibility to react to physicians' behavior, we expect competition to have a positive effect on the consulting and a negative effect on the overtreatment rate (increase of consulting and decrease of overtreatment rate from BASE to COMP). The same considerations apply in the presence of the insurance; hence we expect the consulting rate to increase and the overtreatment rate to decrease from INS to INS-COMP.

In summary, we expect competition (i.e. the free choice of physicians) to reduce overtreatment and thus to increase consulting. That is, we expect the consulting rate to be higher and the overtreatment rate to be lower in COMP than in BASE, and we expect the same to hold in INS-COMP compared to INS.

4. Results

We first present descriptive statistics and discuss aggregated treatment effects. We continue with the analysis of the additional control treatments and a regression analysis. Finally, we

[21] After 10 (15, 20, 30) interactions with a specific physician, a patient needs to receive a severe treatment in at least 7 (9, 11 15) out of these interactions in order to be able to reject the null hypothesis that the physician is always providing the appropriate treatment at the 5 percent level. It is hence rather difficult for patients to detect overtreatment.

present additional health economic measures and discuss cost-effectiveness of alternative institutions in terms of public health status vs. total expenditures, and close with a discussion of inequality amongst patients. In appendix G we describe behavior over time, and in appendix H we discuss whether overtreatment is related to physicians' market shares.

4.1 Descriptive statistics

Table 2 shows the consulting rate (i.e. the share of consulting patients) and the overtreatment rate (i.e. the share of consulted physicians who provide a severe treatment when the problem is mild[22]) averaged across markets and periods in lines (1) and (2). In BASE, both of these shares are at intermediate levels: 40.7 percent of patients consult and only 26.3 percent of consulting patients are being overtreated. Both of these findings are remarkable considering the predictions of standard theory since consulting (i.e. trusting the physician) is ill-advised and the incentives for overtreatment are quite strong. This combination results in a fairly high efficiency rate (realized earnings relative to potential payoff, see line 3).

Table 2: Aggregate results

	BASE	COMP	INS	INS-COMP
(1) consulting rate	40.7	54.7	55.3	83.1
(2) overtreatment rate	26.3	7.2	70.9	34.2
(3) efficiency rate	61.2	70.5	71.5	89.5
(4) correct treatment rate (CTR)	29.6	49.7	16.2	54.9
(5) average earnings physicians	9.1	11.5	14.4	19.1
(6) average earnings patients	6.8	7.2	5.7	6.4

Notes: The table shows averages over all 30 periods and 7 markets in the main treatments. The rates in the first four lines are indicated in percent: (1) is the share of consulting patients, (2) is the share of consulted physicians who give severe treatment when the problem is mild, where the average rate (2) is weighted by the number of consultations per session and period. (3) is the sum of actual earnings over the sum of potential earnings, (4) is the share of all interactions when appropriate treatment is provided. Average earnings in (5) and (6) are indicated in points.

[22] In the remainder of the paper, we present all overtreatment rates conditional on patients having a mild problem (i.e. excluding periods with a severe problem). See appendix C for overtreatment rates including periods with a severe problem which tends to depress the overtreatment rates.

According to the definition of efficiency in (3), overtreatment implies no efficiency loss (overtreatment is a pure redistribution from patients to physicians in our design). According to this measure even a (socially undesirable) situation in which all patients consult and all physicians over-treat is considered efficient. The "correct treatment rate" (CTR) is an alternative measure of efficiency that does not have this property. The CTR is the share of interactions in which the patient gets the treatment he needs. Thus, both the consulting rate and the overtreatment rate determine the CTR.[23] As in the calculation of the overtreatment rate, we consider only periods with a mild problem (for the CTR including periods with a severe problem, see appendix C).[24] The remaining lines (5) and (6) show earnings of physicians and patients, respectively.

4.2 The impact of insurance and competition

Table 2 shows that insurance induces more consulting but also boosts overtreatment, compared to the baseline condition, as expected.[25] The overtreatment rate almost triples (from 26.3 to 70.9 percent) and the correct treatment rate therefore suffers (falls from 29.6 to 16.2 percent), but the consulting rate still increases from 40.7 to 55.3 percent. When using the rather conservative Wilcoxon-Mann-Whitney tests[26] (which treat the average over all periods in market as one independent observation) to assess the significance of these effects, we find that the effect of insurance on the overtreatment rate is highly significant while the effect on consulting is not (see table 3). This lack of significance is perhaps due to the high degree of heterogeneity of the consulting rate across markets in BASE (3 markets have rates around 15 percent, two around 65 percent, see figure D1 in the appendix). It is mostly physicians who benefit from insurance (their incomes increase by more than 50 percent, from 9.1 in BASE to

---

[23] The reason that the CTR is not exactly equal to consulting rate x (1- overtreatment rate) in table 2 is that the trust rate shown in table 2 is the overall consulting rate (share of patients consulting) rather than the trust rate conditional on the problem being mild. The conditional consulting rate is shown in appendix C.

[24] Equivalently, the CTR including periods with severe problem can be calculated using the trust and overtreatment rates including periods with a severe problem (a comparison of all measures presented in table 2 both including and excluding periods with a severe problem can be found in appendix C).

[25] The finding that insurance boosts demand is indeed unsurprising to an economist. Cutler and Zeckhauser (2000) note in the Handbook of Health Economics: "essentially all economists accept that traditional health insurance leads to moderate moral hazard in demand."

[26] The WMW test assumes that the two distributions being tested are identical except for a shift. As this assumption might be violated with experimental data we additionally perform a robust rank order test (see Fligner and Policello 1981) that assumes neither equal variances, nor equal shape of the distributions. The level of significance using the robust rank order test is for no comparison lower, and in some cases higher, than with the WMW test. For the detailed results see appendix C, table C3.

14.4 points in INS on average, and this effect is highly significant, see table 3). Patients suffer as they rush to consult while being overtreated at much higher rates than without insurance (their incomes drop significantly, from 6.8 to 5.7 points. Recall that the costs of being overtreated dominate the benefits of being treated at all). However, the effects of insurance on the CTR are not significant according to the WMW test.

Table 3: Wilcoxon-Mann-Whitney (WMW) test – empirical z-values

	Impact of competition		Impact of insurance	
	Without insurance	With insurance	Without competition	With competition
	BASE vs. COMP	INS vs. INS-COMP	BASE vs. INS	COMP vs. INS-COMP
(1) consulting rate	-1.28	-2.62 ***	-1.09	-2.75 ***
(2) overtreatment rate	3.00 ***	2.36 **	-3.13 ***	-3.07 ***
(3) efficiency rate	-1.34	-2.62 ***	-1.09	-2.75 ***
(4) correct treatment rate (CTR)	-1.60	-2.36 **	0.83	-0.58
(5) average earnings physician	-1.09	-2.49 **	-1.98 **	-2.88 ***
(6) average earnings patients	-1.60	-1.47	2.88 ***	1.73 *

Notes: see table 2 for explanations of variables. Positive numbers indicate that the value of the variable is larger in the treatment condition named first, and vice versa for negative values. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$

Competition in the guise of "informed choice" has largely beneficial effects, as expected: It reduces overtreatment and increases consulting. The effects are rather strong.[27] Overtreatment is cut by about half with insurance (falls from 70.9 to 34.2 percent) and by about two thirds without insurance (from 26.3 to 7.2 percent). Both of these effects are statistically highly significant according to the WMW test, see table 3. Competition increases the consulting rate by about a third absent insurance and by about 50 percent with insurance. The effect of

[27] Recall that we discuss the effects of competition in the guise of "informed choice" in this section, meaning that the comparison of BASE vs. COMP and of INS vs. INS-COMP show the effect of allowing patients to choose a physician *and* of being informed about market shares (i.e. how many patients consult with each physician). The effects of physician choice absent information on market shares (i.e. the effect of "pure competition") and of providing information about market shares are discussed in section 4.3.

competition is highly significant with insurance (i.e. INS vs. INS-COMP) but not without insurance (i.e. BASE vs. COMP) according to a conservative Wilcoxon-Mann-Whitney test.

Perhaps surprisingly, competition does not have much of an effect on patients' earnings (no significant differences).[28] The reason is that the payoff difference from not consulting vs. being treated correctly is rather small (one point payoff difference for a patient). Perhaps even more surprisingly, physicians tend to benefit from competition (the difference is significant for the conditions with insurance). The reason is that the increase in profit from the increase in the consulting rate more than compensates the loss in physicians' profit from the decrease of the overtreatment rate. Regarding efficiency, competition increases both the efficiency rate and the correct treatment rate; the differences are again only significant for the conditions with insurance. The finding that the beneficial effects of competition (in particular on the consulting rate and on the efficiency rate) are stronger with insurance is somewhat surprising. A closer look at the data suggests that this is due to the high variation of the consulting rate across markets in the baseline condition: the average consulting rate is below 20% in three markets while it is higher than 60% in two markets (see appendix D). Therefore, it is likely that the lack of statistical significance comparing the baseline and the condition with competition is due to the rather small number of independent observations per condition.

The effects of insurance given competition are strong, in line with our expectations and significant in all cases (with the exception of the CRT). In particular, insurance induces moral hazard on the side of patients (the consulting rate increases from 54.7 to 83.1 percent), and boosts overtreatment by almost factor 5 (from 7.2 to 34.2 percent). Insurance is again (as in the case without competition) to the benefit of physicians (their incomes increase from 11.5 to 19.1) while patients' incomes even fall (from 7.2 to 6.4 points, weakly significant).

4.3 The impact of market shares

As pointed before, the comparison of the main treatments BASE vs. COMP and INS vs. INS-COMP measures the effect of introducing "informed choice" in the sense that we not only enable patient to choose which physician they want to consult (if any) but at the same time also provide the patients (and the physicians) with information about how popular physicians are by making market shares observable. It is possible that such informed competition is more

[28] Nonetheless, COMP is the only condition in which patients benefit from consulting a physician. In BASE, the expected payoff of consulting (given the observed physician behavior) is lower than the expected payoff of not consulting. In the conditions with insurance, an individual patient should always consult even if he expects to be overtreated (due to the incentives of the insurance, see section 3.2). On the group level, however, the group of all patients would be better off if they refrained from consulting (given the behavior of an average physician).

powerful than “pure competition” without information about market shares as such information might channel patients to popular physicians who, in turn, have sharper incentives to earn popularity by providing adequate treatments. One might therefore wonder what the effects of “pure” competition in isolation are. Conversely, one might wonder what the effects of providing information are in the absence of competition.

To investigate this issue, we implement two control treatments called COMP_nms and INS-COMP_nms (where “nms” stands for “*no market shares*") serving to decompose the total effect of “informed choice” into an effect of “pure competition” and an effect of “market information”. Note that a plausible result of such a decomposition is that none of the two effects turns out significant despite the composite effect being significant. Such a finding would simply point to interaction effects.

Table 4: Descriptive results for the control treatments

	Conditions without insurance			Conditions with Insurance		
	BASE	COMP_nms	COMP	INS	INS-COMP_nms	INS-COMP
(1) consulting rate	40.7	42.2	54.7	55.3	79.8	83.1
(2) overtreatment rate	26.3	20.7	7.2	70.9	44.4	34.2
(3) efficiency rate	61.2	61.8	70.5	71.5	87.7	89.5
(4) correct treatment rate (CTR)	29.6	34.1	49.7	16.2	43.8	54.9
(5) average earnings physicians	9.1	9.2	11.5	14.4	19.1	19.1
(6) average earnings patients	6.8	6.9	7.2	5.7	6.1	6.4
no. markets ; no. subjects	7 ; 56	7 ; 56	7 ; 56	7 ; 56	7 ; 56	7 ; 56

Notes: Table shows averages over all 30 periods and 7 markets per treatment. (1) is the share of patients who consult a physician, (2) is the share of consulted physicians who give severe treatment when the problem is mild, where the average (2) is weighted by the number of consultations per session and period. (3) is the sum of actual earnings over sum of potential earnings, (4) is the share of all interactions with needed treatment provided. Average earnings in (5) and (6) are indicated in points.

The control treatments COMP_nms and INS-COMP_nms are identical to the conditions COMP and INS-COMP, respectively, except for the fact that physicians' market shares are not displayed in the history window of their screens (neither for the patients nor for the physicians). Everything else (in particular the matching procedure) is identical to the conditions with competition.

The follow-up was run at the same laboratory (at the University of Copenhagen) using the same software, subject pool and recruiting system, instructions and procedures as for the main treatments. We made sure that each subject participated in one session only. Overall, 112 subjects participated in seven markets per control treatment. They earned 215.1 DKK each on average which is not significantly different from average earnings in the main treatment (p = 0.631, Wilcoxon-Mann-Whitney tests).

Table 4 shows that the outcomes in the control treatments are without exception between the outcomes in the respective main treatments. That is, every single out of 12 measures for the treatments with competition but without information about market shares is between the respective treatment without such information and the respective treatment with competition. This is a comforting finding, indicating that the results seem fairly robust and can be nicely replicated. Inspection of the values reveals that the values for COMP_nms are closer to BASE than to COMP, and that INS-COMP_nms is closer to INS-COMP. Eyeballing thus indicates that the effect of “pure competition” was stronger in the presence than in the absence of insurance.

Table 5 provides statistical tests for differences between the controls and the respective main treatments that span the total effect of “informed choice”. With one exception, we find that market shares have no independent effect (in addition to the effect of competition) as there are no significant differences between the main treatments involving such information and the respective control (see right half of the table). However, the one exception is somewhat surprising because it indicates (together with the insignificant effect of BASE vs. COMP_nms in line 2) that the total effect of “informed choice” on overtreatment in BASE vs. COMP was mainly driven by information and only secondarily, if at all, by “pure competition”. Yet, from the right half of the table it becomes clear that this is somewhat of an outlier result as all other (11 out of 12) tests indicate non-significance of information. We believe the outlier has to do with the large degree of heterogeneity of markets in BASE and the relatively small number of independent observations (see figure D1).

Table 5: Wilcoxon-Mann-Whitney (WMW) test – empirical z-values

	Effect of pure competition (no market shares)		Effect of information about market shares (in add. to competition)	
	(1) Without insurance	(2) With insurance	(3) Without insurance	(4) With insurance
	BASE vs. COMP_nms	INS vs. INS-COMP_nms	COMP_nms vs. COMP	INS-COMP_nms vs. INS-COMP
(1) consulting rate	-1.19	-2.95 ***	-1.21	-1.41
(2) overtreatment rate	0.06	2.36 **	1.98 **	0.07
(3) efficiency rate	-0.45	-3.00 ***	-1.09	-1.34
(4) correct treatment rate (CTR)	-0.19	-2.75 ***	-1.09	-1.09
(5) average earnings physician	-0.06	-2.36 **	-1.09	-0.19
(6) average earnings patients	-0.32	-1.34	-1.09	-1.09

Notes: see table 2 for explanations of variables. Positive numbers indicate that the value of the variable is larger in the treatment condition named first, and vice versa for negative values. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$

Testing also reveals no evidence of an independent effect of "pure competition" absent insurance (see column 1). But we do find strong evidence of an independent effect of pure competition when insurance is present (see column 2). In fact, 5 out of 6 tests indicate a significant difference (at the 5 percent level or better according to a WMW-test) between outcomes with and without competition when patients are shielded from bearing the full cost of overtreatment (and do not get information about market shares). We summarize the discussion above as follows. In our main treatments, we find strong effects of insurance, and we also find strong effects of competition in the presence, but not in the absence, of insurance (see table 3). However, this effect of competition measures the joint effect of providing information about market shares *and* of allowing patients to choose physicians, i.e. it measures the effect of what we call "informed choice". We have therefore run two control treatments to determine whether that effect is mainly due to the ability to choose as such (i.e. "pure competition" effect) or to the additional information provided. The results in table 5 show that matters are not totally unambiguous in the condition absent insurance (none of the two effect seems to be dominant except for overtreatment which seems to be driven by information provision), but the evidence is clear and unambiguous in the case with insurance. Here we find that the strong effect of “informed competition” was mainly due to “pure competition” and there is no evidence of an independent effect of providing “information” as such. This result is comforting because it upholds our interpretations that differences between the main treatments that do or do not involve competition mainly reflect the effect of competition (“pure and simple”).

4.4 Regression analysis

Table 6 presents the results of a panel probit regression. The second line shows that insurance unambiguously increases both consulting and overtreatment. With respect to competition, we have to decompose the effect of “informed choice” into pure competition and the observability of market shares. For consulting, the regression results indicate that the overall positive effect of competition is to a greater extent driven by free choice than by the observability of market shares (compare line 1 and line 3). For overtreatment, however, pure competition turns out to be insignificant, in contrast to the observability of market shares (the impact of market shares is conditional on the presence of pure competition as we do not conduct a control treatment without free choice but with market shares.

Line 4 shows that time trends are quantitatively weak and tend to be negative for consulting and positive for overtreatment. Line 5 shows that a patient is less likely to consult if he got severe treatments more often than expected (i.e. the dummy Sev_high takes a value of 1 if a patient got severe treatment in more than a third of the cases). Line 6 shows that physicians are more likely to overtreat when the benefits of doing so are high, i.e. when many patients choose to consult with her. All of these effects are robust to the inclusion of gender and field of study as controls.

Lines 7 to 12 show interaction effects. For example, column (C) shows that receiving severe treatments beyond expectations (i.e. when Sev_high = 1) discourages patients from consulting when competition prevails but this effect is weak when insurance spreads the cost of severe treatments to the entire group (-0.18 is more than compensated by 0.26). Line 11 in (C) shows that there are positive interaction effects between insurance and pure competition concerning consultation: adding competition to the conditions with insurance increases the demand for consultation as does adding insurance to the conditions with competition. However, we find no such interaction effects for overtreatment, see line 11 in col. (F). Finally, we find that physicians who are consulted by more than one patient are more prone to overtreat (line 6, col. F), but that this effect is mitigated by competition (see line 9. Appendix H provides more details on this relation).

Table 6: Average marginal effects (panel probit regression)

	Dep. variable	Consulting			Overtreatment		
		(A)	(B)	(C)	(D)	(E)	(F)
1	COMP	0.21***	0.23***	0.23***	-0.10	-0.11	-0.14
2	INS	0.24***	0.24***	0.24***	0.39***	0.35***	0.35***
3	Obs_shares	0.05	0.03	0.04	-0.24***	-0.21***	-0.20**
4	Period	-0.00***	-0.00***	-0.00***	0.00**	0.00**	0.00**
5	Sev_high	-0.20***	-0.20***	-0.18***	-	-	-
6	#Pat_high	-	-	-	0.08***	0.07***	0.06**
7	COMP x Sev_high	-	-	-0.03	-	-	-
8	INS x Sev_high	-	-	0.26***	-	-	-
9	COMP x #Pat_high	-	-	-	-	-	-0.15***
10	INS x #Pat_high	-	-	-	-	-	-0.00
11	COMP x INS	-	-	0.17**	-	-	-0.14
12	Obs_shares x INS			-0.01			-0.00
	Controls	No	Yes	Yes	No	Yes	Yes
	No. of obs.	5346	5346	5346	1632	1632	1632
	AIC	0.958	0.958	0.957	0.914	0.910	0.909

Notes: Table shows average marginal effects from a panel probit regression (see appendix C, table C4 for regression coefficients). Dependent variables are dummies which indicate in (A) to (C) whether or not a patient consults a physician, and in (D) to (F) whether or not a physician overtreats conditional on being consulted. COMP, INS and Obs_shares are dummies indicating the treatment variations. COMP indicates whether patients can choose their physician (i.e. equals 1 for COMP, COMP_nms, INS-COMP and INS-COMP_nms), INS denotes the presence of the insurance, and Obs_shares indicates whether market shares are observable (i.e. equals 1 for COMP and INS-COMP). Sev_high = 1 when the rate of severe treatments a patient got so far exceeds its mathematical expectation, i.e. 1/3, and = 0 otherwise. #Pat_high = 1 if physician was consulted by more than one patient, and = 0 otherwise. Controls "Yes" indicates that the regression includes gender and field of study.

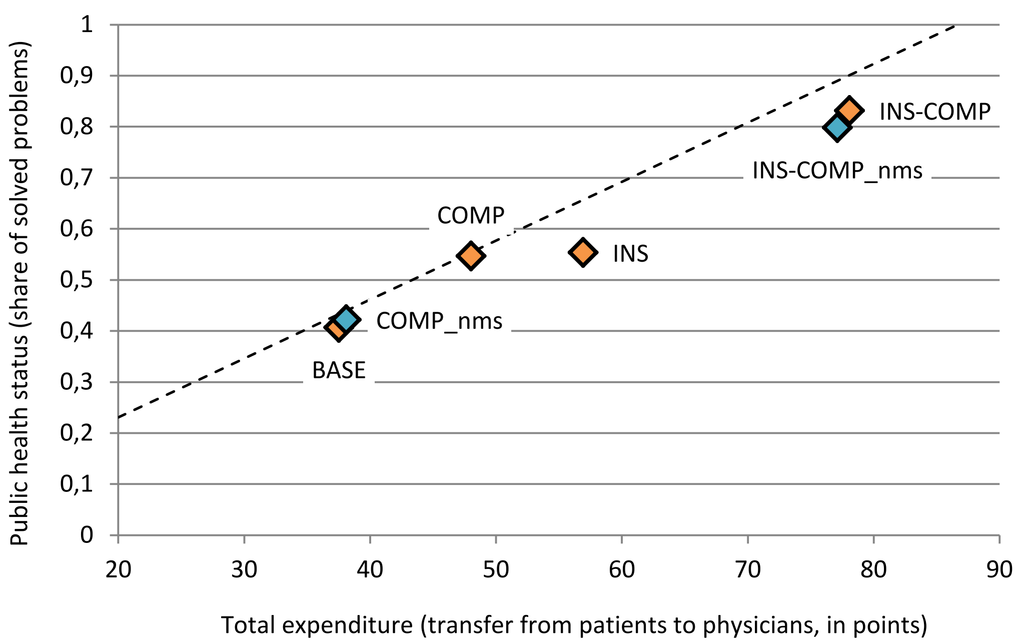
4.5 Measures reflecting public health

This section presents two additional measures serving to evaluate the results of our experiment taking a perspective of public health.

Figure 5 illustrates the cost-effectiveness of alternative institutional arrangements. The figure compares the overall expenditures of the laboratory "healthcare system" under study with the "average health status" in the population.[29] We define the total health expenditures as the average transfer (in points) from the patients to physicians. For a given patient, these expenditures are zero if he does not consult a physician, they are $p(m) = 15$ if he receives a mild treatment and $p(s) = 22$ for a severe treatment. We measure the state of public health by the share of patients whose problem was solved. Note that these two measures are by definition positively associated (by design, a patient's problem can only be solved if a treatment has been provided – and treatments are, again by design, costly). The dashed line indicates the expenditures that are required to reach a particular level of public health assuming there is no overtreatment.

Figure 5 shows that INS and COMP produce a similar level of public health, but INS does so at a cost that is almost 20 percent higher. These excess costs result from the (two-sided) moral hazard effects the insurance coverage induces. Thus, the institution of free choice of physician dominates the institution of insurance in terms of the trade-off shown here if one has to choose between the two.

Figure 5: Public health status vs. total expenditures

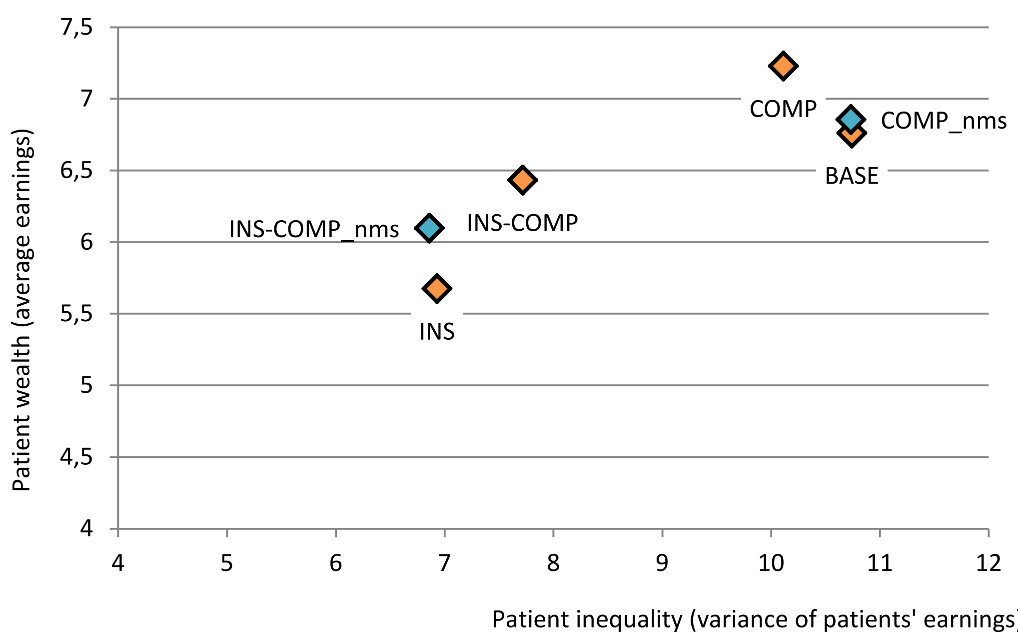


[29] We would like to emphasize that the present analysis refers to the context of our laboratory experiment and that the terms used in this section should not suggest that we are referring to the health care sector in any particular country. The goal of our study is not to give policy recommendations based on our laboratory experiment, but to study human behavior in this particular, rather abstract setting. Of course, the insights of our study have to be corroborated and refined in a field setting before further conclusions should be drawn.

The highest level of public health is achieved in INS-COMP. According to the measure used here, it is about twice the level in BASE. This impressive performance, however, is also associated with the highest total expenditure. The figure thus illustrates the trade-off a society may face when making institutional choices in health care: insurance increases the level of public health substantially by making sure that patients with a problem indeed consult a physician. But insurance at the same time also boost expenditures: the first reason is exactly because the patients want to be treated, the second is because physicians overtreat much more knowing that patients do not have to bear the full cost of overtreatment. Furthermore, the horizontal distance to the dashed line is higher in the conditions with insurance – indicating that the same level could be reached at a lower level of total expenditures.

Figure 6 paints a somewhat more favorable picture of the effects of insurance. While it is true that insurance invites moral hazard from both sides of the market, it also provides the benefit of a more even distribution of (material) welfare among patients. Insurance means that the risk of having to bear the cost of severe treatments is borne by society at large rather than by the individual who has been unfortunate enough to suffer from a severe problem.

Figure 6: The upside of insurance: reduced uncertainty (inequality)



The figure shows that inequality of incomes among patients (measured by the variance of incomes) is lowest when patients are covered by medical insurance and much higher in the conditions without insurance. However, this higher degree of safety (or equality) among patients comes at a price, as patient wealth is also lower with insurance than without.

5. Concluding remarks

We have used the methods of experimental economics to study how competition and insurance shape outcomes in the interaction between patient and physician, with a particular focus on patient overtreatment. The provision of a medical treatment is a typical example of a credence good, and we therefore investigate a stylized market for credence goods.

We find that competition in the guise of giving patients free choice of physician exerts pressure on physicians not to overtreat patients (because overtreating physicians are shunned to some extent). This, in turn, seems to reassure patients and induces additional patients to consult a physician. Additional consultations tend to increase efficiency as more health problems are solved, and solved properly. As expected, we find that insurance induces moral hazard on both sides of the market. With insurance, more patients consult a physician as the additional cost of treatment (including overtreatment) is not borne by the consulting patient but by all patients collectively. The upside of insurance is that it inspires confidence in patients and distributes the cost more evenly among patients. The downside is that physicians respond to more careless demand for medical treatments by overtreating patients more often.

The interaction effects between the institutions are particularly interesting and pronounced. Competition mitigates much of the adverse effects of insurance. Overtreatment rates are not much higher in INS-COMP than in BASE (34 vs. 26 percent, $p = 0.749$ WMW-test). At the same time, the beneficial effect of insurance on inspiring confidence (or trust) in the system is boosted (the consulting rate is much higher, at 83 percent, than in either COMP or INS, both at about 55 percent). As a result, efficiency is high and patients are fairly likely to get the treatment they need. However, physicians reap more of the benefits of having these institutions (their earnings are about three times as high as patients' earnings with insurance, and this is also true in INS-COMP).

From a public health perspective, we identify a trade-off between the average health status in the population and the cost of the healthcare system. We observe the highest level of public health in the condition with insurance and free choice of physician but costs are high and overtreatment is prevalent. It does not seem possible (at least within our simple framework) to reduce costs or the level of overtreatment without decreasing the level of public health at the same time. Comparing the effectiveness of insurance and competition, we find that the level of public health is comparable, but overall expenditures are lower with competition. On the other hand, insurance reduces inequality amongst the group of patients.

We conclude that competition in the guise of free choice of physician has unambiguously beneficial effects in our setting. Informed competition increases the level of public health and decreases overtreatment at the same time. Even though the increased level of public health leads to higher total expenditures, the total expenditures are not as high as they could be – because competition provides incentives to keep overtreatment reasonably low. Of course, extrapolation to health care policies in the field is difficult. Here are three reasons. First, competition is not the only mechanism that can reduce overtreatment, and quite possibly it is not the best that can be thought of. For instance, Brosig-Koch et al. (2014) find that overprovision is significantly lower in a mixed-fee-for-service remuneration system (where the fee-for-service component is complemented by a lump-sum component) compared to a pure fee-for-service system. One would need a comparative evaluation of competition and mixed-fee-for-remuneration or other potential measures to be on safer ground for policy advice. Second, there are distinct differences between the anonymous interactions in an abstract laboratory setting and the interactions of patients and physicians in the field where personal interaction often plays a role. We think it would be interesting for follow-up research to investigate the effects of the institutions studied here in controlled environments that are richer in context. Third, professional norms are also likely to play an important role for the degree to which physicians engage in opportunistic behavior. Kesternich et al. (2015) for instance show (in a controlled laboratory-like internet experiment) that medical students are more likely to sacrifice parts of their own income for a patient's benefit if they are primed for professional norms (in the context of the Hippocratic Oath).

An advantage of our experimental design is that these factors are held constant across conditions. We can hence interpret differences across conditions (instead of absolute levels) and describe the effects of the institutions under investigation. We are able to identify causal mechanisms by implementing ceteris paribus variations which are difficult if not impossible to implement in a field setting. Furthermore, the error-free observability of outcomes – in particular regarding overtreatment – is very difficult to achieve in the field.

Last but not least, it is important to note that our results can only be interpreted in the context of healthcare systems in which physicians benefit from providing a high level of medical care (i.e. in a fee-for-service remuneration system). In healthcare systems where physicians receive a fixed premium per patient (i.e. in a capitation remuneration system), patients tend to suffer from undertreatment rather than overtreatment – a phenomenon we abstract from in our experiment. Our results tentatively suggest that free choice of physician might also be a good instrument to mitigate undertreatment in a capitation remuneration system. A fruitful alley for future research, we think, is to determine whether this is in fact the case.

References

Ahlert, M., Felder, S., Vogt, B., 2012. Which patients do I treat? An experimental study with economists and physicians. Health Economics Review 2(1), 1-11.

Aron-Dine, A., Einav, L., Finkelstein, A., 2013. The RAND health insurance experiment, Three decades later. Journal of Economic Perspectives 27(1), 197-222.

Balafoutas, L., Beck, A., Kerschbamer, R., Sutter, M., 2013. What drives taxi drivers? A field experiment on fraud in a market for credence goods. Review of Economic Studies 80(3), 876-891.

Beck, A., Kerschbamer, R., Qiu, J., Sutter, M., 2013. Shaping beliefs in experimental markets for expert services: Guilt aversion and the impact of promises and money-burning options. Games and Economic Behavior 81, 145-164.

Beck, A., Kerschbamer, R., Qiu, J., Sutter, M., 2014. Car mechanics in the Lab: Investigating the behavior of real experts on experimental markets for credence goods. Journal of Economic Behavior & Organization 108, 166-173..

Berger, L. A., Hershey, J. C., 1994. Moral hazard, risk seeking, and free riding. Journal of Risk and Uncertainty 9(2), 173-186.

Bohnet, I., Harmgart, H., Huck, S., Tyran, J.-R., 2005. Learning Trust. Journal of the European Economic Association 3(3), 322-9.

Bolton, G. E., Katok, E., Ockenfels, A., 2004. How effective are electronic reputation mechanisms? An Experimental Investigation. Management Science 50(11), 1587-1602.

Böhm, R., Betsch, C., Korn, L., 2015. Selfish-rational non-vaccination: Experimental evidence from an interactive vaccination game. Journal of Economic Behavior & Organization (in press), DOI: 10.1016/j.jebo.2015.11.008.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., Wiesen, D., 2013a. How to improve patient care? An analysis of capitation, fee-for service, and mixed payment schemes for physicians. Ruhr Economic Papers, No. 412.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., Wiesen, D., 2013b. How effective are pay-for-performance incentives for physicians? A laboratory experiment. Ruhr Economic Papers, No. 413.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., Wiesen, D., 2015. The effects of introducing mixed payment systems for physicians: Experimental evidence. Health Economics (in press), DOI: 10.1002/hec.3292.

Buckley, N. J., Cuff, K., Hurley, J., McLeod, L., Mestelman, S., Cameron, D., 2012. An experimental investigation of mixed systems of public and private health care finance. Journal of Economic Behavior & Organization 84, 713-729.

Chiappori, P. A., Durand, F., Geoffard, P. Y., 1998. Moral hazard and the demand for physician services: first lessons from a French natural experiment. European Economic Review 42(3-5), 499-511.

Clemens, J., Gottlieb, J. D., 2014. Do physicians' financial incentives affect medical treatment and patient health? American Economic Review 104(4), 1320-1349.

Cooper, Z., Gibbons, S., Jones, S., McGuire, A., 2011. Does hospital competition save lives? evidence from the English NHS patient choice reforms. Economic Journal 121(554), F228-F260.

Cutler, D. M., Zeckhauser, R. J., 2000. The anatomy of health insurance. In: Culyer, A. J., Newhouse, J. P. (Eds.), Handbook of Health Economics, Vol.1. Amsterdam: Elsevier, 563-643.

Darby, M. R., Karni, E., 1973. Free competition and the optimal amount of fraud. Journal of Law and Economics 16(1), 67-88.

Devlin, R. A., Sarma, S., 2008. Do physician remuneration schemes matter? The case of Canadian family physicians. Journal of Health Economics 27(5), 1168-1181.

Dufwenberg, M., Gächter, S., Hennig-Schmidt, H., 2011, The framing of games and the psychology of play. Games and Economic Behavior 73, 459-478.

Dulleck, U., Kerschbamer, R., 2006. On doctor, mechanics, and computer specialists: The economics of credence goods. Journal of Economic Literature 44(1), 5-42.

Dulleck, U., Kerschbamer, R., Sutter, M., 2011. The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. American Economic Review 101(2), 526-555.

Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10(2), 171-178.

Fligner, M. A., Policello, G. E., 1981. Robust rank procedures for the Behrens-Fisher problem. Journal of the American Statistical Association 76(373), 162-168.

Green, E. P., 2014. Payment systems in the healthcare industry: An experimental study of physician incentives. Journal of Economic Behavior & Organization 106, 367-378.

Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. An online recruitment system for economic experiments. Journal of the Economic Science Association 1(1), 114-125.

Hennig-Schmidt, H., Wiesen, D., 2014. Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students. Social Science and Medicine 108, 156-165.

Hennig-Schmidt, H., Selten, R., Wiesen, D., 2011. How payment systems affect physicians' provision behavior — an experimental investigation. Journal of Health Economics 30(4), 637-646.

Huck, S., Lünser, G. K., Tyran, J.-R., 2012. Competition fosters trust. Games and Economic Behavior 76(1), 195-209.

Huck, S., Lünser, G. K., Tyran, J.-R., 2016. Price competition and reputation in markets for experience goods: An experimental study. RAND Journal of Economics 47(1), 99-117.

Huck, S., Normann, H.-T., Oechssler, J., 2004. Two are few and four are many: number effects in experimental oligopolies. Journal of Economic Behavior & Organization 53(4), 435-446.

Iizuka, T., 2007. Experts' agency problems: evidence from the prescription drug market in Japan. RAND Journal of Economics 38(3), 844-862.

Kairies, N., Krieger, M., 2013. How do non-monetary performance incentives for physicians affect the quality of medical care? A laboratory experiment. Ruhr Economic Papers, No. 414.

Kalda, R., Polluste, K., Lember, M., 2003. Patient satisfaction with care is associated with personal choice of physician. Health Policy 64(1), 55-62.

Keser, C., Montmarquette, C., Schmidt, M., Schnitzler, C., 2013. Custom-made healthcare – An experimental investigation. CIRANO-Scientific Publications, No. 2013s-15.

Keser, C., Peterle, E., Schnitzler, C., 2014. Money talks: Paying physicians for performance. CIRANO-Scientific Publications, No. 2014s-41.

Kesternich, I., Schumacher, H., Winter, J., 2015. Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus? Journal of Public Economics 131, 1-11.

Ledyard, J. O., 1995. Public Goods. A survey of experimental research. In: Kagel, J. H., Roth, A. E. (Eds.). Handbook of Experimental Economics. Princeton, NJ: Princeton University Press, 111-194.

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., Leibowitz, A., 1987. Health insurance and the demand for medical care: evidence from a randomized experiment. American Economic Review 77(3), 251-277.

Mimra, W., Rasch, A., Waibel, C., 2013. Price competition and reputation in credence goods markets: Experimental evidence. CER-ETH Zurich Working Paper No. 13/176.

Mimra, W., Rasch, A., Waibel, C., 2014. Second opinions in markets for expert services: Experimental evidence. CER-ETH Zurich. Working Paper No. 14/192.

The Insurance Experiment Group, 1993. Free for all?: lessons from the RAND health insurance experiment. Cambridge, MA: Harvard University Press.

Riedl, A., Tyran, J.-R, 2005. Tax liability side equivalence in gift-exchange labor markets. Journal of Public Economics 89(11), 2369-2382.

Schmittdiel, J., Selby, J. V., Grumbach, K., Quesenberry, C. P., 1997. Choice of a personal physician and patient satisfaction in a health maintenance organization. Journal of the American Medical Association 278(19), 1596-1599.

Schneider, H. S,, 2012. Agency problems and reputation in expert services: evidence from auto repair. Journal of Industrial Economics 60(3), 406-433

Schram, A., Sonnemans, J., 2011. How individuals choose health insurance: an experimental analysis. European Economic Review 55, 799-819.

Scitovsky, A. A., Snyder, N. M., 1972. Effect of coinsurance on use of physician services. Social Security Bulletin 35, 3-19.

Sørensen, R. J., Grytten, J., 2003. Service production and contract choice in primary physician services. Health Policy 66(1), 73-93.

Sülzle, K., Wambach, A., 2005. Insurance in a market for credence goods. Journal of Risk and Insurance 72(1), 159-176.

Appendix A – Instructions for treatment INS-COMP

(*Instructions for the other treatments are available from the authors on request*)

Welcome to the experiment.

Please read these instructions carefully. Do not speak to the other participants and keep quiet during the entire experiment. In case you have a question please raise your hand. We will then come to you.

At the beginning of the experiment you are randomly assigned to **groups of 8 participants – out of which 5 will be Clients and 3 will be Advisers**. You will be informed whether you are a Client or an Adviser at the beginning of the experiment. During the experiment you solely interact with the participants of your group.

In this experiment you can earn money. The show-up fee amounts to 75 DKr for both Clients and Advisers. During the experiment we do not talk about DKr, but about points. You can earn additional points according to the choices you make. These points will be converted into Danish Crowns (DKr) and paid to you in cash at the end of the experiment according to the following exchange rate:

2 points = 1 DKr

How much you earn depends on your decisions and on the decisions of other participants in your group. All participants receive the same instructions. All decisions are made anonymously. That is, no other participant will get to know your name or your income.

General Description

You are in a group with 7 other participants. At the beginning of the experiment all participants are randomly assigned to one of two roles (**Client** or **Adviser**). You will be informed about your role at the beginning of the experiment. There are five Clients and three Advisers (*A1, A2, and A3*). All participants keep their role and the number assigned to them throughout the experiment.

The experiment consists of **30 periods**. In each period, the Clients have a new **problem**. All the Clients have the same problem. In each period the computer randomly determines whether the problem is **severe** or **mild**. The problem is **severe** in **one third** of the cases, and in **two thirds** of the cases the problem is **mild**:

1/3 of the cases the problem is **severe**
2/3 of the cases the problem is **mild**

In each period, the Clients have to decide if they want to consult an Adviser or not. If they want to consult, they have to choose which Adviser to consult. The Clients **do not know** whether their problem is severe or mild. **The Advisors, however, do know the severity of the clients' problem.** If a Client consults an Adviser, the Adviser provides a treatment that solves the problem. If a Client decides not to consult an Adviser, the problem will not be solved. Hence, if a

Client decides to consult an Adviser, the Client's problem will always be solved, but he will not learn the true severity of the problem. As a result, the Adviser may provide the **severe** treatment, even though the true severity of the problem is only **mild**.

After each period, the participants will be informed of their income in this period and of their total income earned so far. We explain how incomes are calculated below.

Task of a Client

At the beginning of each period, each Client has to decide whether or not to consult an Adviser. This means that a Client has to decide whether he wants to consult A1, A2, A3, or none of them.

At the beginning of a period, the computer randomly determines the true severity of the problem. All Advisers are informed about the true severity of the problem, but the Clients **do not** know the true severity of the problem. If a Client chooses not to let an Adviser solve the problem, he will learn the true severity of the problem at the end of the period. If he chooses to consult an Adviser, he will not learn the true severity of the problem, but is only informed about the treatment provided by the Adviser.

Task of an Adviser

The Advisers know the true severity of the Clients' problems regardless of whether or not any of the Clients chose to consult him. As there are a total of 5 Clients, an Advisor can be consulted by up to 5 Clients in one period. An Adviser only learns the number of Clients who chose to consult him, but he will not learn the identity of those who chose to consult him. If an Adviser is consulted by any of the Clients, the Adviser has to provide the same treatment to all Clients. If the problem is **severe** he has to provide the **severe** treatment. If the problem is **mild**, he can provide either the **mild** treatment or the **severe** treatment.

Income

The Income Table (see separate sheet) shows the incomes for both an Adviser and a Client for all possible cases.

The Income earned by a Client:

If a Client chooses not to consult an Adviser, he gets no treatment. In this case, his income depends on the severity of his problem and on the total number of severe treatments in the group. If his problem is **severe** he earns $2-\frac{\#\cdot of\ \cdot severe\cdot treatments}{5}*7$ points (where "*# of severe treatments*" is short for "number of severe treatments") or 0 if the amount in this formula becomes negative (Nobody can never make a loss in this experiment.). If, on the other hand, his problem is **mild** he earns $9-\frac{\#\cdot of\ \cdot severe\cdot treatments}{5}*7$ points. See first line in the Income Table.

If a Client decides to consult an Adviser, his income will not depend on the treatment provided by the Adviser, but solely on the total number of severe treatments in the group. If the Adviser decides to provide the mild or the severe treatment, the client earns $10-\frac{\#\cdot of\ \cdot severe\cdot treatments}{5}*7$ points.

Notice, however, that while the Advisor will always provide the **severe** treatment when the problem is **severe**, he can decide to provide either the **severe** treatment or the **mild** treatment when the problem is **mild**. Clients who consult an Advisor will always learn if they received the mild or the severe treatment. Moreover, all Clients will observe the total number of severe treatments in the group (Advisor's don't).

The income earned by an Advisor:
Remember that it is possible for an Adviser to be chosen by **several** Clients. This means that the income earned by an Adviser from the interaction with one Client, as shown in the Income Table, should be **multiplied** by the number of clients that have chosen to consult him.

Examples:

Suppose the true severity of the problem is **mild**, an Adviser has been chosen by 4 Clients, and the Adviser chooses to provide the **severe** treatment. In this case the Adviser's income is (4*17) = 68. The fifth Client does not consult an Advisor.
The income of the 4 Clients who sought treatment is 10 – (4/5)*7 = 4.4 while the income of the fifth Client who did not seek treatment is 9 – (4/5)*7 = 3.4.

Suppose the problem is **mild** and two Clients have decided to consult an Advisor. We see from the Income Table that the income earned by the Adviser from one Client is 10 points if he provides the **mild** treatment while it is 17 points if he provides the **severe** treatment. As two Clients decided to consult him, his total income in this period is (2*10) = 20 points if he provides the **mild** treatment while it is (2*17) = 34 points if he provides the **severe** treatment.
Only one other Client consults a different Advisor and gets a **severe** treatment.
In the first case where the first two Clients get a mild treatment *all* three Clients who sought treatment earn 10 – (1/5)*7 = 8.6 each while the income of those who didn't seek treatment is 9 – (1/5)*7 = 7.6. In the second case with severe treatment for the first two Clients *all* three Clients who sought treatment earn 10 – (3/5)*7 = 5.8 each while the income of those who didn't seek treatment is 9 – (3/5)*7 = 4.8.
If the problem was **severe** all Clients who sought treatment would earn 10 – (3/5)*7 = 5.8 and the Clients who didn't seek treatment would earn 0 (because 2-(3/5)*7=-2.2 is negative). Remember that all consulted Advisors have to provide the severe treatment if the problem is severe.
If an Adviser is not consulted in a period, his income from that period is 0 points.

The History Table

In order for the participants to keep track of what has happened in previous periods, both the Advisers and the Clients are shown history tables on the left part of their screen. These tables are mere summaries. They do not provide information in addition to what has been told during the experiment, with **one exception**: The History Table for both Clients and Advisers shows the number of Clients that chose to consult a particular Adviser in all past periods.

These are the rules. You can trust us that everything will happen exactly according to these rules. Take your time to go over the instructions once again and feel free to ask questions. But don't shout! Simply raise your hand.

Income Table

		True severity of problem			
		Severe		Mild	
Adviser provides	No Treatment	$2 - \frac{\#\cdot of \cdot severe \cdot treatments \cdot}{5} * 7$ or 0 if negative	0	$9 - \frac{\#\cdot of \cdot severe \cdot treatments \cdot}{5} * 7$	0
	Severe Treatment	$10 - \frac{\#\cdot of \cdot severe \cdot treatments \cdot}{5} * 7$	17	$10 - \frac{\#\cdot of \cdot severe \cdot treatments \cdot}{5} * 7$	17
	Mild Treatment	-	-	$10 - \frac{\#\cdot of \cdot severe \cdot treatments \cdot}{5} * 7$	10
		Client Income	Advisor Income	Client Income	Advisor Income

Note: The table shows the income for an Advisor from being consulted by **one** Client

Appendix B – Screenshots

Figure B1: Screenshot physician (adviser) – condition BASE and INS

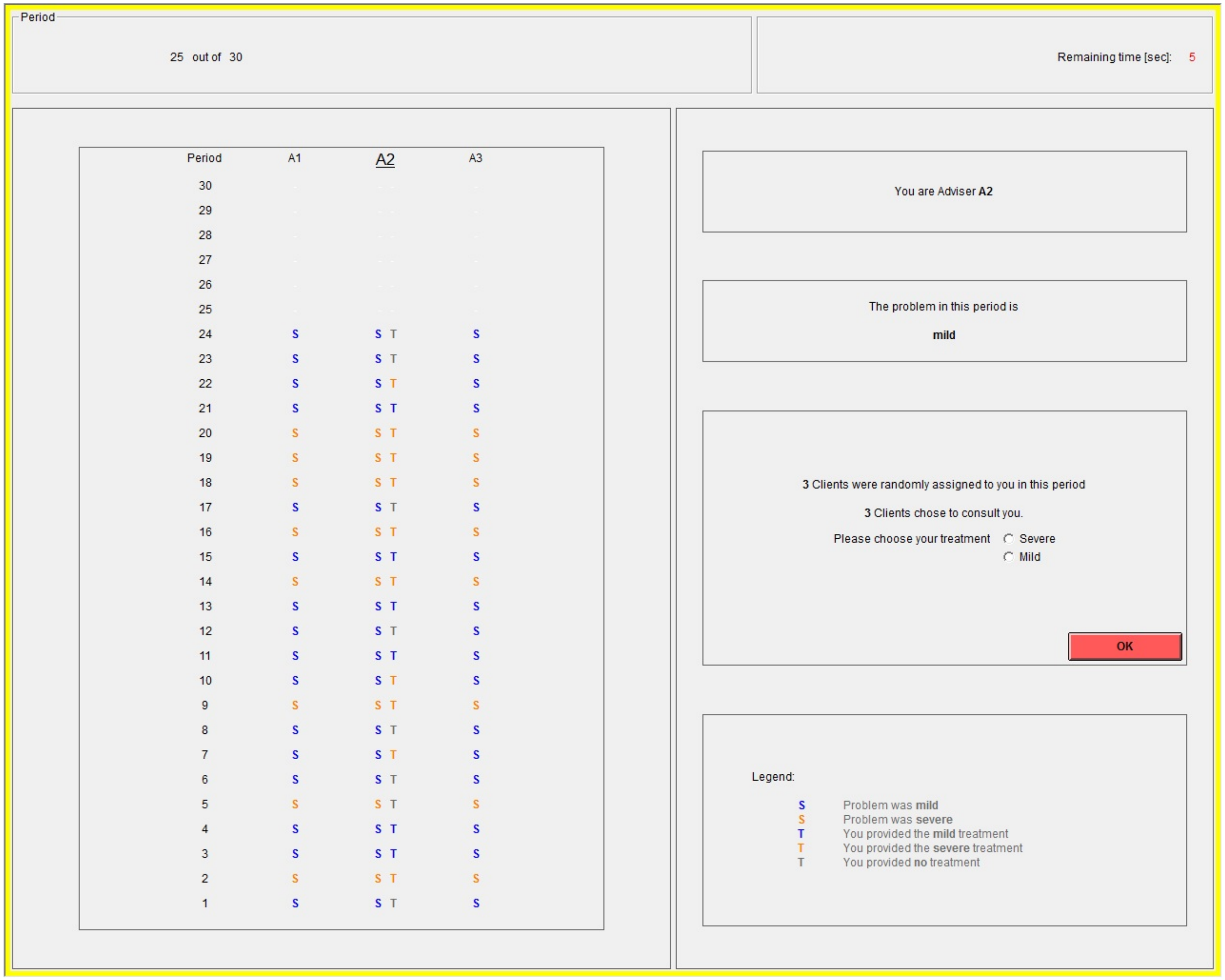


Figure B2: Screenshot physician (adviser) – condition COMP and INS-COMP

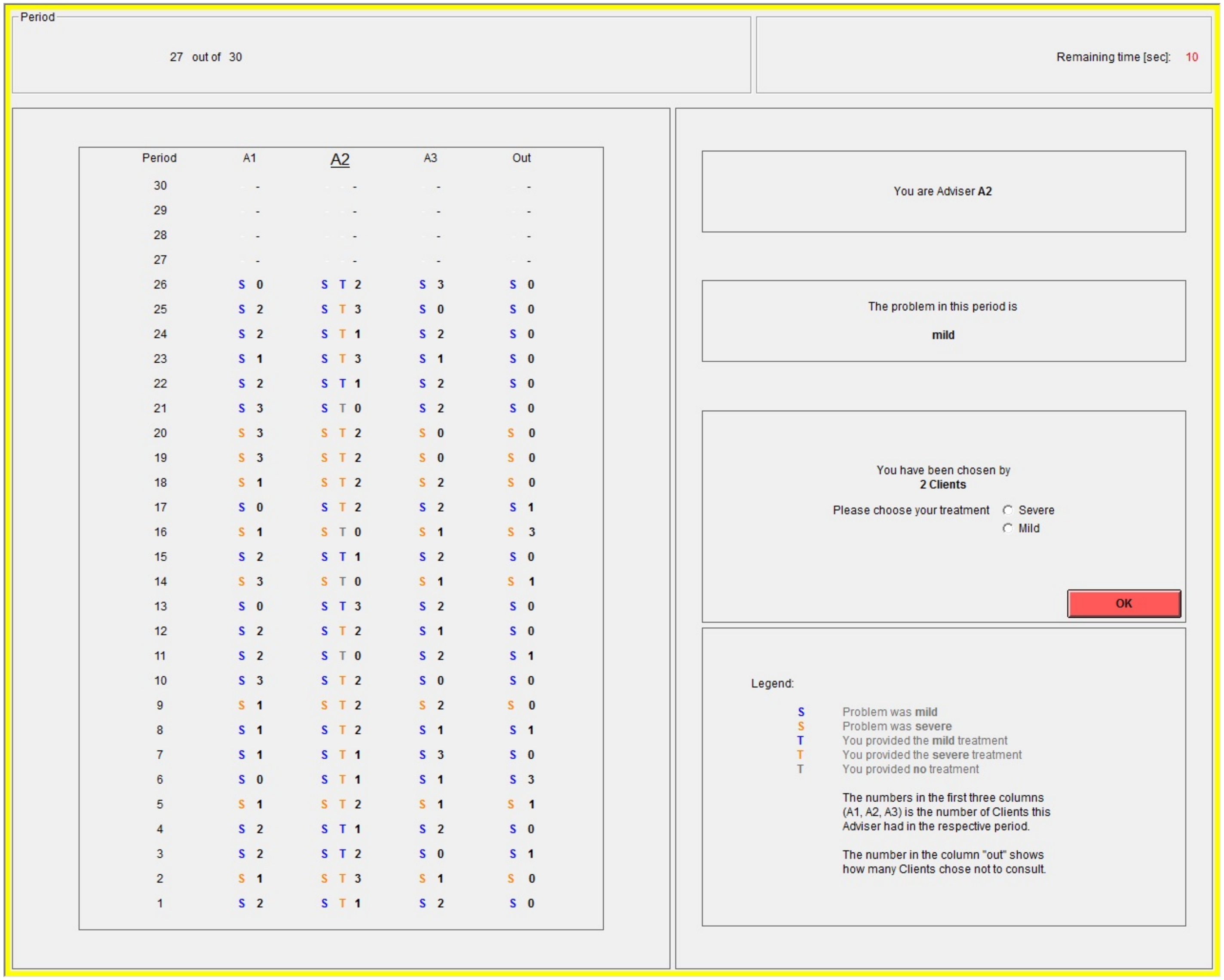


Figure B3: Screenshot patient (client) – condition BASE and INS

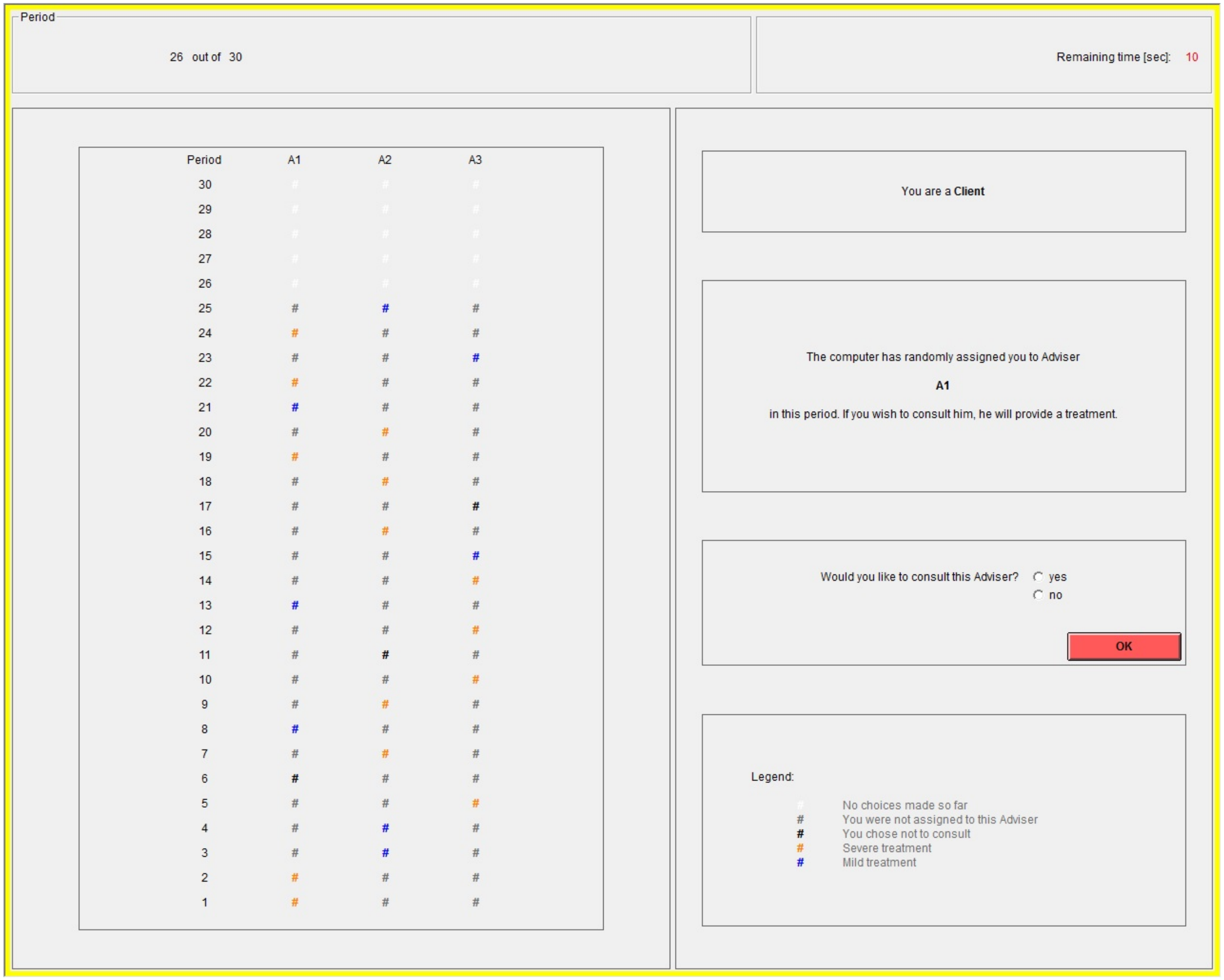
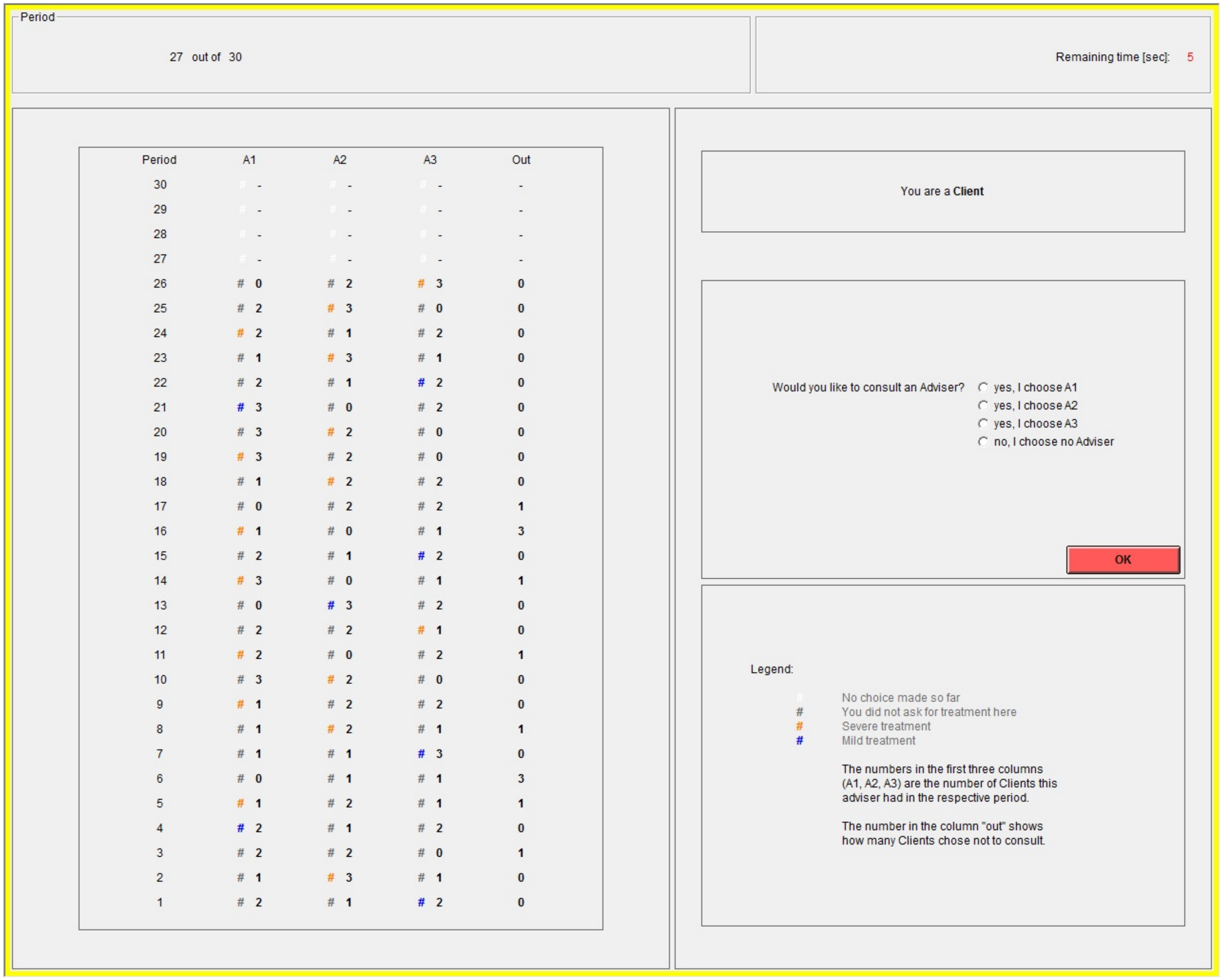


Figure B4: Screenshot patient (client) – condition COMP and INS-COMP



Appendix C – Tables including / excluding severe periods

Table C1: Aggregated results (including and excluding severe periods)

	Severe periods	BASE	COMP	INS	INS-COMP
(1) consulting rate	included	40.67	54.67	55.33	83.14
	excluded	40.14	53.56	55.48	83.42
(2) overtreatment rate	included	18.03	4.88	49.40	23.83
	excluded	26.28	7.16	70.86	34.15
(3) efficiency rate	included	61.17	70.50	71.47	89.46
	excluded	67.08	74.46	75.51	90.88
(4) correct treatment rate (CTR)	included	33.34	52.00	28.00	63.34
	excluded	29.59	49.73	16.16	54.93
(5) average earnings physicians	included	9.12	11.46	14.37	19.10
	excluded	7.92	9.37	13.83	17.23
(6) average earnings patients	included	6.76	7.23	5.67	6.43
	excluded	8.66	9.27	6.80	7.84

Notes: Table shows averages over all 30 periods and 7 groups per treatment. The rates in the first eight lines are indicated in percent: (1) is the share of consulting patients, (2) is the share of consulted physicians who give severe treatment when the problem is mild, where the average rate (2) is weighted by the number of consultations per session and period. (3) is sum of actual earnings over sum of potential earnings, (4) is the share of all interactions with needed treatment provided. Average earnings in (5) and (6) are indicated in points. The data for "included" is reprinted from table 2 in the main text.

Table C2: Wilcoxon-Mann-Whitney test – emp. z-values (including / excluding sv. periods)

		Impact of competition		Impact of insurance	
		Without insurance	With insurance	Without competition	With competition
	Severe periods	BASE vs. COMP	INS vs. INS-COMP	BASE vs. INS	COMP vs. INS-COMP
(1) consulting rate	included	-1.28	-2.62 ***	-1.09	-2.75 ***
	excluded	-1.09	-2.62 ***	-1.15	-2.75 ***
(2) overtreatment rate	included	2.88 ***	2.36 **	-3.13 ***	-3.13 ***
	excluded	3.00 ***	2.36 **	- 3.13 ***	-3.07 ***
(3) efficiency rate	included	-1.34	-2.62 ***	-1.09	-2.75 ***
	excluded	-1.09	-2.62 ***	-1.15	-2.75 ***
(4) correct treatment rate (CTR)	included	-1.34	-2.50 **	0.70	-0.96
	excluded	-1.60	-2.36 **	0.83	-0.58
(5) average earnings physicians	included	-1.09	-2.49 **	-1.98 **	-2.88 ***
	excluded	-0.58	-1.85 *	-2.234 **	-3.13 ***
(6) average earnings patients	included	-1.60	-1.47	2.88 ***	1.73 *
	excluded	-3.00 ***	-1.34	3.13 ***	2.88 ***

Note: see table C1 for explanations. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$

Table C3: Robust rank order test – Empirical Û-values

	Impact of competition		Impact of insurance	
	Without insurance	With insurance	Without competition	With competition
	BASE vs. COMP	INS vs. INS-COMP	BASE vs. INS	COMP vs. INS-COMP
(1) consulting rate	-1.32	-4.63 ***	-1.05	-6.02 ***
(2) overtreatment rate	14.32 ***	3.04 **	n. d. ***	-24.98 ***
(3) efficiency rate	-1.40 *	-4.69 ***	-1.05	-6.14 ***
(4) correct treatment rate (CTR)	-1.81 **	-3.04 **	0.72	-0.53
(5) average earnings physicians	-1.05	-4.22 ***	-2.51 **	-8.64 ***
(6) average earnings patients	-1.81 **	-1.44 *	8.64 ***	1.99 **

Note: If the highest observation in one condition is smaller than the lowest observation in the other treatment, the test statistic of the robust rank order test is not defined. We denote these cases with n. d. and three stars. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$

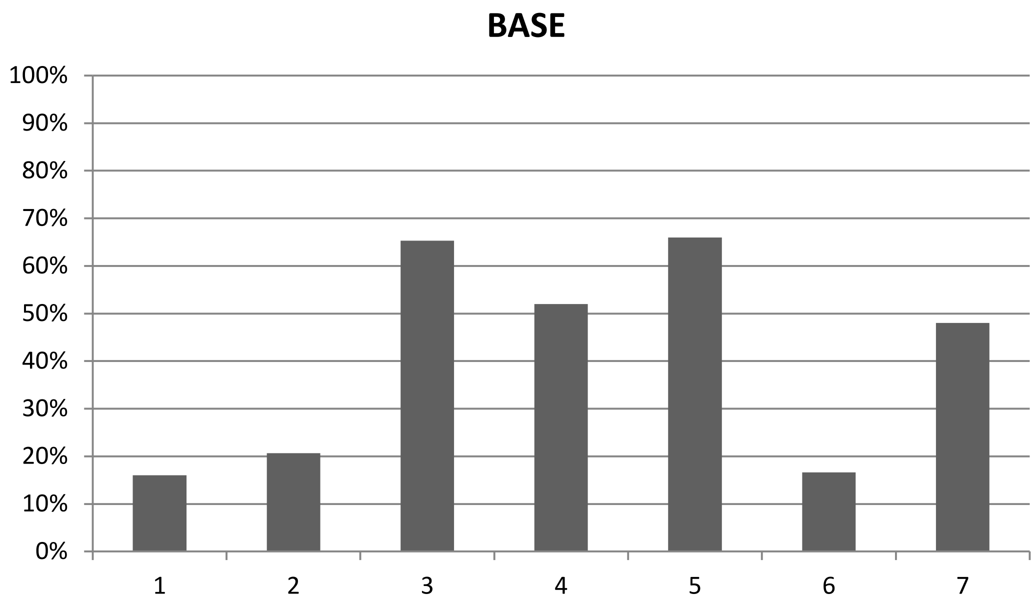
Table C4: Regression coefficients (panel probit regression)

	Dep. variable	Consulting			Overtreatment		
		(A)	(B)	(C)	(D)	(E)	(F)
1	COMP	0.66***	0.74***	0.67**	-0.34	-0.38	0.04
2	INS	0.78***	0.80***	-0.20	1.21***	1.19***	1.44***
	Obs_shares	0.16	0.12	0.12	-0.80***	-0.75**	-0.84**
3	Period	-0.02***	-0.02***	-0.02***	0.01**	0.01**	0.01***
4	Sev_high	-0.82***	-0.81***	-0.76***	-	-	-
5	#Pat_high	-	-	-	0.27***	0.27***	0.62***
6	COMP x Sev_high	-	-	-0.34	-	-	-
7	INS x Sev_high	-	-	0.63***	-	-	-
8	COMP x #Pat_high	-	-	-	-	-	-0.54***
9	INS x #Pat_high	-	-	-	-	-	-0.00
10	COMP x INS	-	-	0.67**	-	-	-0.42
	Obs_shares x INS	-	-	0.01			0.18
11	Intercept	0.66***	0.32	0.56*	-0.75***	-1.56***	-1.76***
	Controls	No	Yes	Yes	No	Yes	Yes
	No. of obs.	5346	5346	5346	1632	1632	1632
	AIC	0.958	0.958	0.957	0.914	0.910	0.909

Notes: Table shows regression coefficients from a panel probit regression. Dependent variables are dummies which indicate in (A) to (C) whether or not a patient consults a physician, and in (D) to (F) whether or not a physician overtreats conditional on being consulted. COMP, INS and Obs_shares are dummies indicating the treatment variations. COMP indicates whether patients can choose their physician (i.e. equals 1 for COMP, COMP_nms, INS-COMP and INS-COMP_nms), INS denotes the presence of the insurance, and Obs_shares indicates whether market shares are observable (i.e. equals 1 for COMP and INS-COMP). Sev_high = 1 when the rate of severe treatments a patient got so far exceeds its mathematical expectation, i.e. 1/3, and = 0 if it is smaller. #Pat_high = 1 if the physician was consulted by more than one patient, and = 0 otherwise (i.e. if she was consulted by one patient only). Controls “Yes” indicates that the regression includes gender and field of study.

Appendix D – Consulting rate in BASE by market

Figure D1: Heterogenous consulting rates across markets



Appendix E – Derivation of the equilibrium in INS

We solve the one-shot game by backwards induction. As in the baseline condition, in the last move, physicians provide the severe treatment, i.e. physicians' incentives do not change with insurance. The decision problem of patient i is as follows: Denote by $\sigma_{-i} \in \{1, 2, 3, 4\}$ the number of other patients in the market who consult. As physicians always provide the severe treatment in equilibrium, σ_{-i} is the total number of severe treatments if patient i does not consult. His expected payoff if he does not consult a physician is:

$$E\pi_i(\sigma_{-i}, \neg C) = \frac{1}{3} \cdot \left(2 - \frac{\sigma_{-i}}{5} \cdot 7\right) + \frac{2}{3} \cdot \left(9 - \frac{\sigma_{-i}}{5} \cdot 7\right) = \frac{20}{3} - \frac{\sigma_{-i}}{5} \cdot 7$$

The expected payoff of patient i if he does consult the physician is calculated analogously. Now, $\sigma_{-i} + 1$ is the total number of severe treatments. As the costs of a severe treatment are socialized, the (expected) payoff does no longer depend on the severity of the problem:

$$E\pi_i(\sigma_{-i}, C) = 10 - \frac{\sigma_{-i} + 1}{5} \cdot 7 = 10 - \left(\frac{\sigma_{-i}}{5} \cdot 7 + \frac{7}{5}\right) = \frac{43}{5} - \frac{\sigma_{-i}}{5} \cdot 7$$

As patient i's (expected) payoff for consulting is greater than the expected payoff for not consulting for all $\sigma_{-i} \in \{1, 2,3, 4\}$, it is dominant to consult. The same holds true for all other patients in the market. Therefore, a unique equilibrium prevails in which patients consult the assigned physician and physicians provide a severe treatment.

Appendix F – Supplementary test results for the control treatments

Table F1: Wilcoxon-Mann-Whitney (WMW) test – empirical z-values

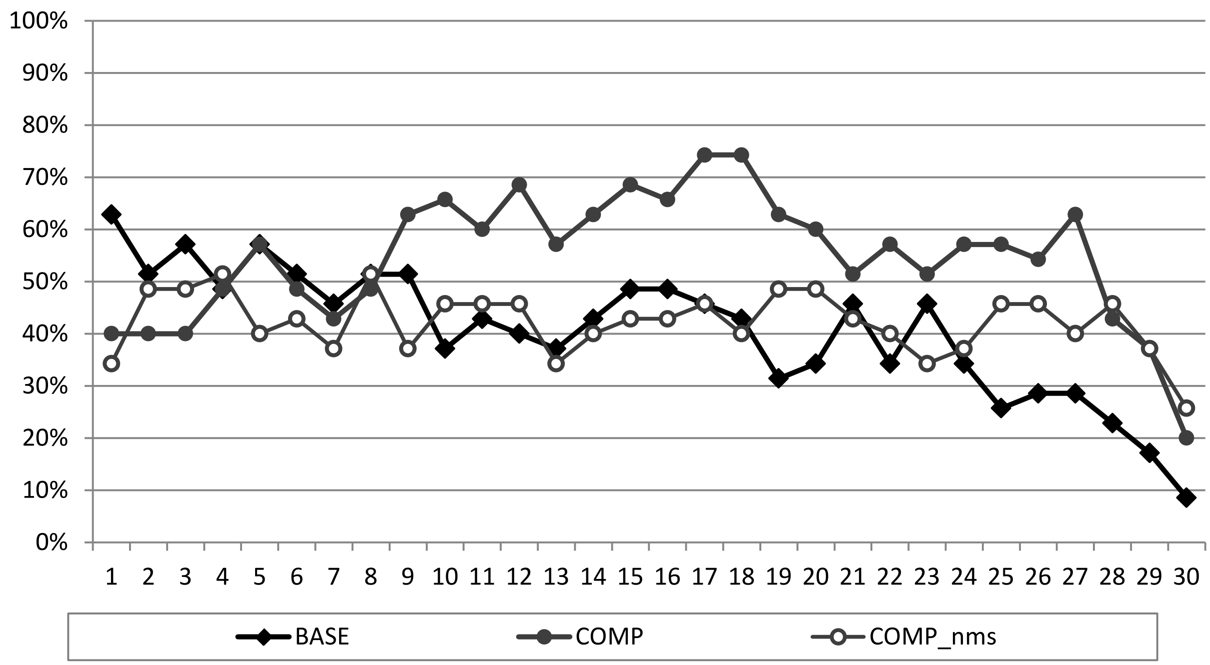
	Impact of insurance		
	(1) Without competition	(2) With competition but without market shares	(3) With competition and market shares
	BASE vs. INS	COMP_nms vs. INS-COMP_nms	COMP vs. INS-COMP
(1) consulting rate	-1.09	-3.01***	-2.75 ***
(2) overtreatment rate	-3.13 ***	-0.96	-3.07 ***
(3) efficiency rate	-1.09	-3.00***	-2.75 ***
(4) correct treatment rate (CTR)	0.83	-0.70	-0.58
(5) average earnings physician	-1.98 **	-3.13***	-2.88 ***
(6) average earnings patients	2.88 ***	1.85*	1.73 *

Notes: see table 2 in the main text for explanations of variables. Positive numbers indicate that the value of the variable is larger in the treatment condition named first, and vice versa for negative values. Numbers in column (1) and (3) reprint columns (3) and (4) from table 3 in the main text.
* $p \le 0.1$; ** $p \le 0.05$; *** $p \le 0.01$

Appendix G – Market outcomes over time

Figure G1 shows the average consulting rate over time in conditions without insurance; figure G2 in the conditions with insurance. In BASE, the consulting rate clearly decreases over time. It is at 57.1 percent in the first three periods and declines to less than a third of that (16.2) percent in the last three. As patients cannot unambiguously tell whether they have been overtreated due to the very nature of the credence good studied here (they can only compare the number of severe treatments they got with the probability of a severe treatment), they have to learn that consulting is not well-advised from experience, and such learning takes time.

Figure G1: Consulting rate over time (conditions without insurance)



In COMP the consulting rate has an inverse u-shape. It starts at around 40 percent in the first 3 periods, peaks at around 70 percent in the middle periods, and falls to around 30 percent in the last three periods. The decline in the second half seems to be driven by the physicians' diminishing incentive to maintain a good reputation over time.

In the treatments with insurance, the consulting rate seems to have no clear trend. It fluctuates between 40% and 60% in INS and remains at considerably higher levels of between 70% and 90% in INS-COMP. This finding is in line with our expectation that both insurance and competition have a positive impact on consulting. In all conditions (except for INS-COMP_nms) we observe pronounced end game effects as reputational concerns wither.

Figure G2: Consulting rate over time (conditions with insurance)

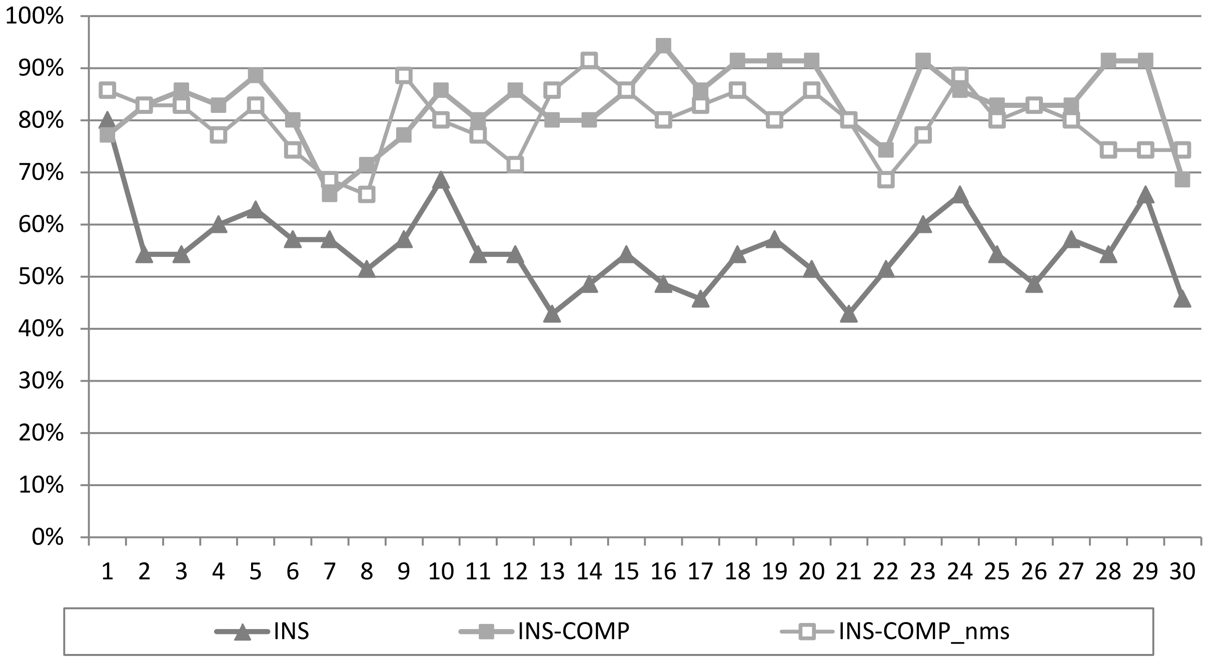
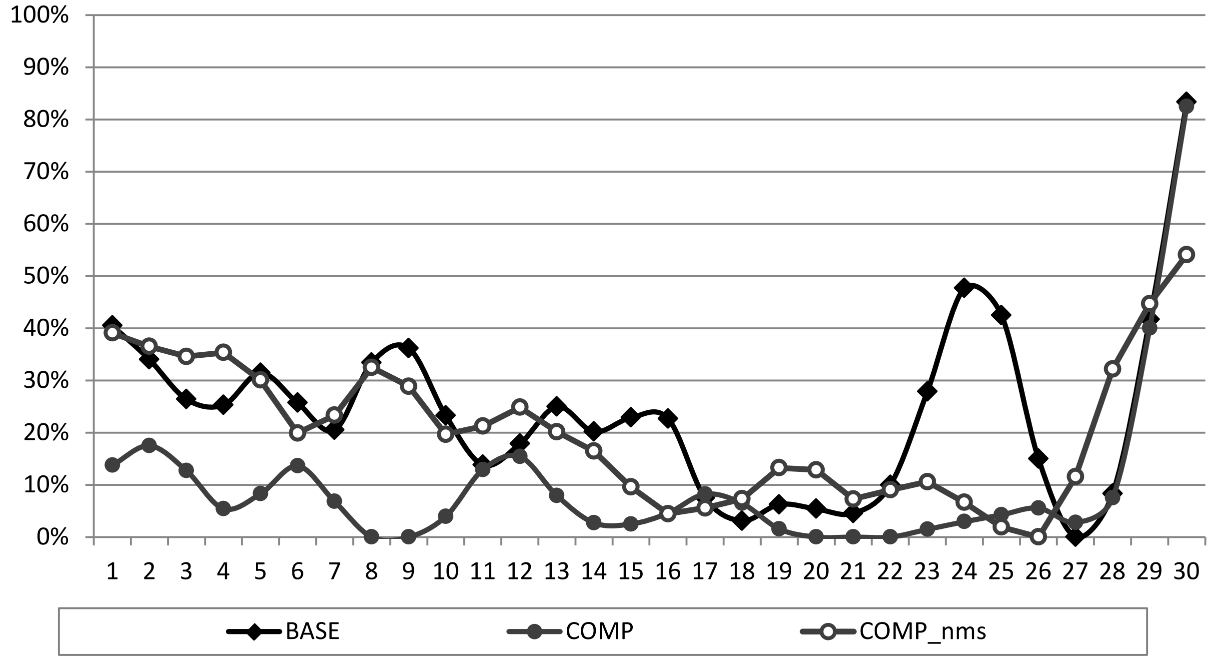


Figure G3 and G4 show the smoothed overtreatment rate over time.[30] The figures shows that competition massively reduces overtreatment, especially in the presence of insurance (compare INS vs. INS-COMP). In COMP, overtreatment is below 20 percent in all periods (except for the pronounced endgame effect), and falls to levels close to zero in most periods in the second half of the experiment. Remarkably, the effect of competition is immediate (absent insurance): the overtreatment rate is lower in COMP than in the other treatments already in the first period. This suggests that physicians anticipate the competitive pressure resulting from the free choice of physicians from the very beginning. The figure also shows the rather detrimental effect of insurance on overtreatment. After some periods and with few exceptions later on, the overtreatment rate hovers around 60 to 90 percent in INS (see line with triangles). Physicians seem to quickly learn that patients have an incentive to consult and do not seem to care much about being overtreated. The negative effect of insurance (second degree moral hazard) seems to slightly dominate the beneficial effect of competition by disciplining service providers in our setting (overtreatment in INS-COMP tends to be slightly higher than in BASE).

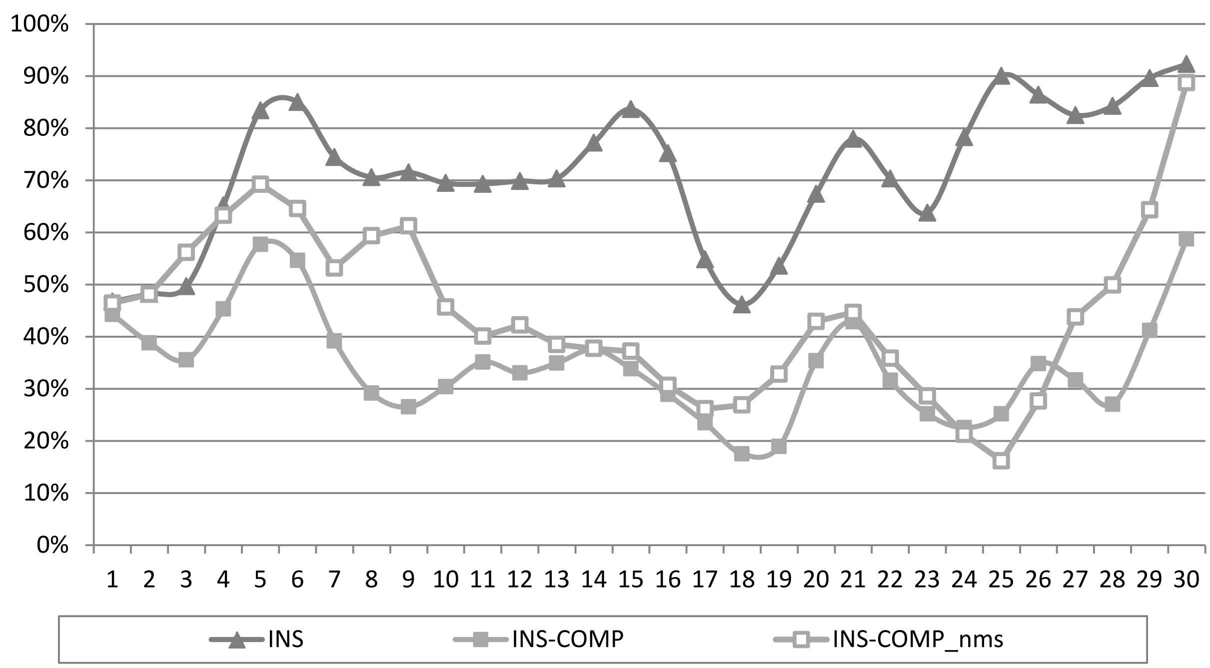
[30] We show smoothed values for the overtreatment rate here to facilitate readability of the figure. This rate fluctuates much more than the consulting rate over time, especially in the conditions with low consulting rates.

Figure G3: Overtreatment rate over time (conditions without insurance)



Notes: Figure shows smoothed overtreatment rate (moving average with a weight of 0.25 on the preceding and the consecutive period, except for periods 1 and 30)

Figure G4: Overtreatment rate over time (conditions with insurance)



Notes: Figure shows smoothed overtreatment rate (moving average with a weight of 0.25 on the preceding and the consecutive period, except for periods 1 and 30)

Appendix H – Overtreatment rates by physician market share

Table H1 addresses the question of whether more popular physicians are less likely to behave opportunistically or, conversely, whether patients tend to consult those physicians who are less prone to behave opportunistically (our analysis cannot distinguish between the two explanations). Specifically, the table shows average overtreatment rates by the number of consulting patients. For example, in BASE, 18.3 percent of all physicians with exactly one consulting patient overtreat that patient. The column marked "overall" recapitulates the numbers from the second line in table 2 in the main text.

Table H1: Overtreatment rate (in percent) by number of consulting patients

	Overall	Number of consulting patients		
		one	two	three or more
BASE	26.3	18.3	27.2	40.9
COMP_nms	20.7	30.4	16.4	9.5
COMP	7.2	9.3	7.3	0.0
INS	70.9	65.9	74.4	75.0
INS-COMP_nms	44.4	57.2	48.0	29.0
INS-COMP	34.2	43.4	31.8	29.8

Notes: the share of "unemployed" physicians, i.e. the share of physicians with zero patients was 53.7 in BASE, 51.60 in COMP_nms, 37.7 in COMP, 36.5 in INS, 28.77 in INS-COMP_nms and 19.2 in INS-COMP (see table E3 for details). The column marked "overall" shows overtreatment rates conditional on being consulted, i.e. it shows the share of consulted physicians who give severe treatment when the problem is mild.

The table illustrates the disciplining effect of competition on physicians. In the conditions without competition, the overtreatment rate of physicians with two, and of those with three and more patients, are higher than the overall overtreatment rate. For example, the overtreatment rate among physicians with two patients consulting in INS is more than twice the rate in INS-COMP (74.4 vs. 31.8 percent), and the overtreatment rate in BASE it is about four times the rate in COMP (27.2 vs. 7.3 percent). The differences are even more pronounced in the (rare) cases where a physician has three or more patients consulting.[31] These results are consistent with our findings from in table 6. These regressions show that overtreatment is significantly lower in the treatments with competition (see line 1). It is generally higher among physicians that are consulted by many patients (see line 6), but that this effect is significantly reduced in the presence of competition (see col. F, line 9).

We find that the "market concentration" (i.e. the concentration of patients on particular physicians) tends to be larger in the treatments with competition. This suggests that patients succeed in selecting physicians who provide more adequate treatments. For example, the share of physicians with two or more consulting agents is about 50 percent larger in COMP than in BASE (22.8 vs. 15.1 percent), and it is almost twice as large in INS-COMP as in INS (44.5 vs. 23.3 percent).

Table H2: Distribution of physicians's market shares

	Number of consulting patients			
	zero	one	two	three or more
BASE	53.65	31.28	10.05	5.02
COMP_nms	51.60	30.82	12.56	5.02
COMP	37.67	39.50	18.72	4.11
INS	36.53	40.18	17.81	5.48
INS-COMP_nms	28.77	31.51	23.29	16.44
INS-COMP	19.18	36.30	33.79	10.73

Notes: The theoretical benchmark of random assignment (if all patients consult the physician they have been assigned to) is 13.17 (zero patients), 32.92 (one patient), 32.92 (two patients) and 20.99 (three or more patients).

[31] This is the case in 4 to 6 percent of the periods in all conditions except for INS-COMP where physicians have three or more patients consulting in about 10 percent of the periods. For details on the distribution of physicians' market shares, see table H2.

Discussion Papers of the Research Area Markets and Choice 2014

Research Unit: **Market Behavior**

Hidekazu Anno, Morimitsu Kurino Second-best incentive compatible allocation rules for multiple-type indivisible objects	SP II 2014-201
Pablo Guillen, Rustamdjan Hakimov Monkey see, monkey do: truth-telling in matching algorithms and the manipulation of others	SPII 2014-202
Roel van Veldhuizen, Joep Sonnemans Nonrenewable resources, strategic behavior and the Hotelling rule: An experiment	SPII 2014-203
Roel van Veldhuizen, Hessel Oosterbeek, Joep Sonnemans Peers at work: From the field to the lab	SPII 2014-204
Sebastian Kodritsch On Time-Inconsistency in Bargaining	SPII 2014-205
Dietmar Fehr, Julia Schmid Exclusion in the All-Pay Auction: An Experimental Investigation	SPII 2014-206
David Danz The curse of knowledge increases self-selection into competition: Experimental evidence	SPII 2014-207
Isa E. Hafalir, Rustamdjan Hakimov, Dorothea Kübler, Morimitsu Kurino College Admissions with Entrance Exams: Centralized versus Decentralized	SPII 2014-208

Research Unit: **Economics of Change**

Jan Grohn, Steffen Huck, Justin Mattias Valasek A note on empathy in games	SP II 2014-301
Maja Adena Tax-price elasticity of charitable donations - Evidence from the German taxpayer panel	SP II 2014-302
Orazio Attanasio, Britta Augsburg, Ralph De Haas, Emla Fitzsimons, Heike Harmgart Group lending or individual lending? Evidence from a randomized field experiment in Mongolia	SP II 2014-303
Britta Augsburg, Ralph De Haas, Heike Harmgart and Costas Meghir Microfinance at the margin: Experimental evidence from Bosnia and Herzegovina	SP II 2014-304

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice

Andrew Schotter, Isabel Trevino SP II 2014-305
Is response time predictive of choice? An experimental study of threshold strategies

Lukas Wenner SP II 2014-306
Expected Prices as Reference Points - Theory and Experiments

Steffen Huck, Gabriele Lünser, Florian Spitzer, Jean-Robert Tyran SP II 2014-307
Medical Insurance and Free Choice of Physician Shape Patient Overtreatment. A Laboratory Experiment

WZB Junior Research Group: **Risk and Development**

Ferdinand M. Vieider, Abebe Beyene, Randall Bluffstone, Sahan Dissanayake, Zenebe Gebreegziabher, Peter Martinsson, Alemu Mekonnen SP II 2014-401
Measuring risk preferences in rural Ethiopia: Risk tolerance and exogenous income proxies

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice